

3次元音空間の高忠実レンダリング技法の研究

著者	千葉 武尊
学位授与機関	Tohoku University
URL	http://hdl.handle.net/10097/39901

東北大学審査修士学位論文

3次元音空間の 高忠実レンダリング技法の研究

2009年2月10日

東北大学大学院情報科学研究科

システム情報科学専攻

千葉 武尊

緒言

人々は楽園のような場所で平和な日々を過ごしていた。

ところがある時主人公は何者かに襲われ、そのまま意識を失ってしまう。
気が付くと、そこはヘッドマウントディスプレイを装着したおびただしい数の人々が、
口々に支配者を崇^{あが}めている異様な空間。
そこは、偽^{いつわ}りの楽園世界に浸る、世界の真の姿であった。

これは、私が中学生の頃流行していたある映画のワンシーンである。

このシーンに登場するシステムは、ヘッドマウントディスプレイを通して作られた映像を見せ、作られた音を聞かせることで、その人間に、あたかも自分がその空間にいるかのような感覚を与える「バーチャルリアリティ技術」を応用したものと考えられる。結局そのシステムはラストシーンで主人公に破壊されてしまうのだが、当時の私はこれに強い興味を覚えた。どうすればこんなことが実現できるのだろうと散々悩んだ挙句、こんなものできるわけがないという結論に至り、少しだけ悔しい思いをしたことを覚えている。

しかし、その後の技術の進歩は私の想像をはるかに超えるものであった。現在、自分がまさにその場にいるかのような感覚、いわゆる「臨場感」を与えるための技術として、このバーチャルリアリティ技術が盛んに開発されている。

バーチャルリアリティ (Virtual Reality) という言葉を聞いたことのある人は多いだろう。この言葉は、1987 年に NASA (アメリカ航空宇宙局) が VPL Research 社に発注して開発した「VIEW」(バーチャル環境ワークステーション) というシステムの開発プロジェクトに際して使い始めた語で、語感の先進的な響きとシンプルさが受け入れられ、コンピュータシステムで現実感を作り出す技術の総称として広く定着している。

バーチャルリアリティの「バーチャル」は、「仮想」とか「虚構」あるいは「擬似」と日本語に訳されることが多い。しかし、この訳では、バーチャルリアリティを「現実世界を模擬することによって形成される人工的な空間」ととらえていることになってしまう。このため、バーチャルリアリティという言葉は虚構であるとか、擬似など

といったネガティブな観念としてとらえられてしまうことが少なくない。

日本バーチャルリアリティ学会初代会長の舘氏によると、このとらえ方が生じてしまう理由は、日本においてバーチャルに対応する正しい概念が存在しなかったためだという。言い換えれば、正しい意味での「バーチャル」にあたる言葉は日本語の中には存在しないのである。「バーチャル」は、The American Heritage Dictionary (21st Century Reference) によれば、“Existing in essence or effect though not in actual fact or form” と定義されている。つまり、「実際の形はしていない、あるいは実物そのものではない、しかし体験としては実際と同じものである、あるいは本質的には実際と同じである」。これがバーチャルという単語の正しい意味である。したがって、日本語における仮想とはまったく逆の意味となる。このような背景から、前述の舘氏は第10回バーチャルリアリティ学会の際に、「バーチャル」を表す漢字として「慣」を提案している（立心偏に「実」の正字体である「實」）。実物のエッセンス、機能としてそのものであるという意味を「心」と「実」を合わせた漢字によって見事に表現されている。バーチャルリアリティは現実そのものではないけれども、現実の本質を持っているのである。

すなわち、普段の生活においても、視覚や聴覚といった五感を通じて、我々が物質そのものの持つ本質を、自分の脳内で昇華することを考えると、極論的な言い方をすれば、我々が実生活で感じるすべての感覚そのものがバーチャルリアリティであるということもできる。このように考えると、バーチャルリアリティは科学技術と共に発展してきた人間生活をより豊かにするための「本質」を、人間の感覚へ映し出すための技術であるにとらえることができる。

バーチャルリアリティの分野で一般的に言われる「臨場感」という感覚は、特に視覚、聴覚を対象として古くから研究がなされており、特に視聴覚コンテンツの再生において常に追求されてきた感覚である。我々が視聴覚コンテンツを鑑賞する際、その作品の世界へ自分が登場しているかのような感覚、没入感を求め、よりリアルにその世界を体感することを望むためである。そして現在、これらの没入感を感じられるような「場」を提示するための技術として、バーチャルリアリティ技術は目覚ましい成長を遂げている。

一例として、視覚について考えてみると、現在の視聴覚コンテンツのほとんどには、従来のフィルム技術に加え、デジタル画像処理や3次元グラフィックスをはじめとするコンピュータグラフィックス技術が用いられており、従来の2次元的な映像表現に加え、視点が動的に変化するなどの3次元的な映像表現が可能となった。また、聴覚においても、従来の2チャンネルスピークを用いたステレオ再生方式から、5.1チャンネルサラウンドシステムに代表されるマルチチャンネル再生方式が一般的になってきており、左右方向のみの表現から、前後左右の平面に自由度が増加した表現がなされるよ

うになり、より実際の環境に近い音空間を表現することが可能になった。また、それらは決して映画館などの特別な施設のみに限定されたものではなく、ホームシアターシステムの普及などにより、一般家庭の視聴覚システムでも気軽に導入がなされ、多くの人が容易に体験できるものとなっている。このように、我々のバーチャルリアリティに対する注目度は明らかに高まってきており、更なる高性能化、更なる再現度向上の実現が期待されていることは容易に考えることができる。

このような背景を考えると、バーチャルリアリティ技術の発展は、情報コンテンツをはじめとするコンテンツ産業の中で重要な位置を占めていることがわかる。特に、コンテンツ制作のための環境の充実は、非常に重要な項目であることは明らかであり、事実、様々なコンテンツ制作ツールの開発が行われている。しかし、視覚に対してはデジタル画像編集ツールや、3次元モデリングソフト等の開発が盛んに行われているのに対し、聴覚に対してはこのようなツールの開発はあまり行われていない。それどころか、聴覚に対しては基本的な3次元音響技術の制作ツールの開発さえ十分に行われていない状況である。現在広く普及しているドルビー 5.1 チャンネルサラウンドシステムは、前後左右といった水平面の表現までしか行うことができず、上下方向からの音の到来を含め、十分に3次元位置を表現しているとは言い難い。加えて、製作者が自分の表現したい音空間を一から形成し、その意図どおりに聴取者に聞かせるためには、これらのハードウェア技術のほかに、3次元音空間をレンダリング(演出)する技術が必要となってくる。例えば、音楽ホールのような広い部屋で音を聞いている場合は、狭い部屋で音を聞いている場合に比べて音がよく響いて聞こえ、同様に、音源と聴取者の間に障害物がある場合、何もない場合と比べてこもったような音になることは経験的に明らかであろう。ところが、現在の技術でこれらの音を再現しようとした場合、大まかな音響効果を付加して「なんとなくそれらしく聞こえる音」にするしかない。アニメーションや日本語吹き替えされた映画で、登場人物の声が不自然にはっきり聞こえたり、逆に響きが大きすぎることに違和感を感じたことのある人は少なくないであろう。

また、現在の3次元音響システムでは、ある特定の音のみを対象として音響処理を行っていることも、忠実性を高めるための障害となっていると考えられる。我々が耳にする実際の音空間においては、特定の音のみが聞こえるという状況は存在せず、室内においては換気扇やエアコンから発せられる音、屋外においては風や雑踏の音などの、背景として存在する様々な音が存在するのが一般的である。ところが、多くの場合は、目的とする特定の音の表現のみを対象としており、背景として存在する音の表現については十分な考慮がされていない。これは既存の3次元音響表現技術が一つの音を表現することに特化した手法であるために、複数の、実際は無数に存在するはずの背景音源を表現することを考慮していないためである。これは室内空間、屋外

空間を問わず、実際の音空間の再現のためには必然的に考慮しなければならない点である。このような状況を表現しうる 3 次元音響効果の付加アルゴリズムは未だに提案されておらず、また、さらなる忠実性の向上といった観点からも、背景音を表現可能なアルゴリズムの提案は重要であると考えられる。

このように、我々の見聞きしている環境に忠実なコンテンツの充実を考えると、3 次元音響技術の整備は非常に重要な要素であり、一刻も早い実現が必要であると考えられる。したがって著者はこのような状況に注目し、本研究を行おうと思うに至った。

本研究では、以上のように 3 次元音響技術の更なる忠実性向上の実現のためのアルゴリズムの構築を目的とする。ここで言う「忠実性」とは、音が物理的に正確であることだけを意味しているわけではなく、聴取者が実際に音を聞いた時に、現在のシステムの音に比べて「より現実の音に近い」と感じる音を「忠実性の高い音」と定義する。したがって音場再現の正確さを重視する物理的アプローチによる検討に加え、人間が実際に聞いた場合の知覚変化を重視する知覚的アプローチを含めた検討を行う。本研究の構成は以下の通りである。まず第 1 章は序論とし、本研究の背景を含め 3 次元音響技術について概説する。第 2 章では、音が回り込んで伝わる現象である回折音に焦点を当て、先行研究から回折音場の最適レンダリングアルゴリズムを検証し、アルゴリズムの高速化手法及び実装法について論述する。第 3 章、第 4 章では、聞かせることを目的とする音ではなく、その背景として存在する音に焦点を当て、背景音が音の感じ方に与える影響や、効果的なレンダリング手法の提案を行う。第 5 章は結論とし、まとめと今後の展望に関して述べる。

目次

緒言	i
第1章 序論	1
1.1 研究背景	1
1.2 3次元音空間のレンダリング方法の現状と課題	2
1.2.1 頭部伝達関数と聴覚ディスプレイシステム	2
1.2.2 音響アプリケーション制作ミドルウェア (SiFASo)	3
1.2.3 回折音の表現	4
1.2.4 背景音の表現	6
1.3 研究の目的	7
第2章 回折音のレンダリングに関する検討	11
2.1 はじめに	11
2.2 Huygens の原理	11
2.3 回折音場のレンダリングに関する検討	13
2.3.1 過去に提案された回折音場のレンダリング手法	13
2.3.2 本研究で採用する回折音レンダリング手法の詳細	18
2.4 SiFASo への実装	20
2.5 レンダリングアルゴリズムの評価	22
2.5.1 測定に使用するラウドスピーカについて	22
2.5.2 回折音場の測定とシミュレーション結果の比較	25
2.6 まとめ	26
第3章 背景音のレンダリングに関する検討	29
3.1 はじめに	29
3.2 背景音に関する先行研究	29
3.3 背景音レンダリング手法の検討	30
3.3.1 背景音に使用する音源について	30
3.3.2 レンダリング手法の検討	33
3.4 主観評価に基づくレンダリングアルゴリズムの評価	35

3.5	まとめ	44
第4章	頭部運動を考慮した背景音のレンダリング	47
4.1	はじめに	47
4.1.1	耳入力音信号の動的変化の効果	47
4.1.2	頭部運動と背景音についての仮説	48
4.2	実験 2-1：頭部運動と背景音の関連についての検証	49
4.3	実験 2-2：背景音群の振幅に変化を与えた場合における検証	54
4.4	実験 2-3：背景音源の音圧レベル分布をパラメータとした検証	57
4.5	まとめ	61
第5章	結論	63
付録A	3章の実験結果の詳細	67
A.1	実験 1-1	67
A.2	実験 1-2	71
付録B	4章の実験 2-3 結果の詳細	75
	謝辞	81
	参考文献	83

第1章 序論

1.1 研究背景

現在，ソフトウェアやゲームといった情報コンテンツは盛んにその開発が進められており，コンテンツ産業は，数ある産業のなかでも重要な位置を占めている．また，我が国においてもコンテンツ産業は経済波及効果が期待される産業分野として位置づけられている．特に，マルチメディアコンテンツに注目してみると，近年その発展は非常に目覚ましいものがあり，我が国においては，アニメーションや映画，ソフトウェアといった分野で大きく注目されている．このような現状を考えた場合，これらのコンテンツの制作環境の整備は，我が国の産業の発展のために非常に重要な課題である．

近年，このようなコンテンツの作成において，コンテンツそのものの内容に加えて，3次元空間のレンダリング技術も注目されている．視覚効果においては，多くのコンテンツで3次元コンピュータグラフィックスが用いられ，音響効果においては，ドルビー 5.1 チャンネルサラウンド方式が一般家庭にも普及するようになってきている．また，最近では3次元ホログラフィを用いた裸眼立体映像再生 [1] や，NHK のスーパーハイビジョンでの 22.2 チャンネルサラウンド方式 [2] に代表されるマルチチャンネルサラウンド方式による立体音響効果のための音響再生装置が開発されている．

ところが，3次元レンダリング技術は，画像等の視覚効果において技術やツール（例えば [3, 4]）等の整備が盛んに行われているのに対し，音響効果におけるコンテンツ作りの基盤技術はほとんど整備されていない．例えば，ドルビー 5.1 チャンネルサラウンド方式に注目してみると，これらは，前後左右といった平面方向における音響効果のみしかレンダリングできず，全方向の包囲感といった観点からは十分な効果が得ることができない．この方式以上の音響レンダリング手法に関しては，バイノーラル録音・再生 [5, 6] のような，様々な技術提案がなされてはいるものの，未だコンテンツ制作環境としての提供は実現されていない．

以上のような背景を考えると，バーチャルリアリティ技術における高精度な3次元音響技術の整備は，情報コンテンツ産業分野の進歩及び発展に大きく貢献することができると考えられ，より高品質なコンテンツの実現を目指す場合，全方向から音が放射される音響効果は，画像などによる視覚効果との相乗効果により，豊かな印象が

得られることが期待できる．そこで，本研究では音響効果に注目したバーチャルリアリティ技術の提供について検討していくことにした．

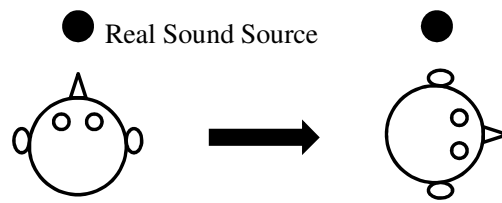
1.2 3次元音空間のレンダリング方法の現状と課題

1.2.1 頭部伝達関数と聴覚ディスプレイシステム

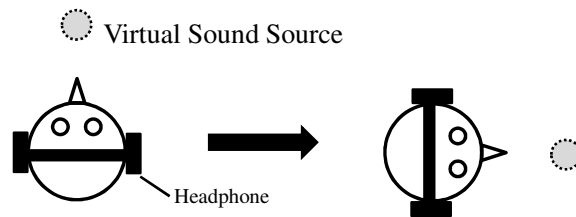
我々は，両耳に到達する音の差異や周波数スペクトルをもとにして，音源までの距離や方向といった空間的な情報を取得することができる．この能力，すなわち音像定位能力を利用し，音源の位置によって耳元での音情報が変化するさまを人工的に再現することで，様々な位置にある音像を仮想的に出現させることが可能である．頭部伝達関数 (Head-Related Transfer Function: HRTF)[7] は，反射のない自由空間内における音源から聴取者の両耳までの音の経路を表す伝達関数であり，音像定位の手がかりを内包した関数である．以下では，特に断らない限り，頭部伝達関数は東北大学電気通信研究所無響室において，球状スピーカアレイを用いて測定されたもの [8] を使用する．

この頭部伝達関数に対応したインパルス応答である，頭部インパルス応答 (Head-Related Impulse Response: HRIR) を測定し，時間軸上で音源に畳み込むことにより，で音像位置の提示が実現される．この原理に基づくシステムが聴覚ディスプレイシステム (Virtual Auditory Display: VAD) である．原理上，聴覚ディスプレイの提示デバイスはラウドスピーカでもヘッドホンでも構わないが，クロストークがなく，聴取時に周囲の音の影響を受けにくいことから，ヘッドホンによる提示法が広く用いられている．しかしながら，通常のヘッドホン聴取時において再生される音信号は，頭部の動きによらず一定であり，頭部を動かすと音像も一緒に動いてしまうため，聴取者は音像の絶対位置を特定することができず，結果として提示される聴覚空間の忠実性は大きく損なわれてしまう (図 1.1(b)) ．

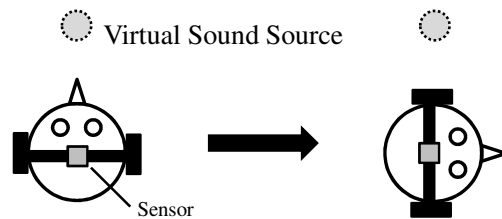
実環境における音像定位では，聴取者の動きや音源の移動などにより生じる音信号の動的な変化が定位精度の改善に大きく寄与することが知られている．このため，聴覚ディスプレイでは聴取者の動きに対応する音像の相対的な運動への対応が重要であり，頭部の動きを検出して，それに応じて畳み込む頭部インパルス応答を切り替えることで，実音源聴取時 (図 1.1(a)) を模擬可能な頭部運動感応型の聴覚ディスプレイ (図 1.1(c)) が実現されている [9, 10, 11, 12, 13] ．代表的な聴覚ディスプレイを，表 1.1 に示す．



(a) 実音源聴取時の様子



(b) センサのない聴覚ディスプレイ



(c) 頭部運動感応型聴覚ディスプレイ

図 1.1: 頭部運動と音源位置の関係

1.2.2 音響アプリケーション制作ミドルウェア (SiFASo)

このような背景から，音響アプリケーション制作ミドルウェア SiFASo(Simulative environment For Acoustic 3D Software) が豊田ら [9] によって開発された．豊田らは，実装に際して，提案アルゴリズムの具体的な実現方法，及び規模に関する検討を行っており，特に室内効果のレンダリング方法について，詳細なフィルタ構成及び設計方法に関する検討を行っている．また，このシステムは，Microsoft 社の Microsoft Windows 上の DLL(Dynamic Link Library) として構築され，オブジェクト指向プログラミングを用いることによって，拡張性，柔軟性，メンテナンス性に富んだシステムとして構築されている．さらに，アプリケーションの自由度を増加させることを目的に，頭部伝達関数による音の到来方向の表現や，初期反射音などによる室内

空間音響の再現など，様々なレンダリング種類を用意し，制作するアプリケーションに応じた選択が可能となっている．特に，移動音像の表現の実現は，既存のシステム [10, 14, 15, 16, 17, 18] には見られない画期的な特徴を有している．アプリケーションはこの DLL を使って，レンダリング種類に応じたパラメータを一定の更新間隔で指定するだけで 3 次元音響出力が得られるため，高度なレンダリング技術を完全にカプセル化して利用できる．また，SiFASo は 1 つの音オブジェクトに対し 1 つのスレッドが生成されるというマルチスレッドの構造をとっており，それぞれのスレッドの出力が，別のミキシングスレッドによって合成されるため，それぞれのスレッドの独立性が強い．

SiFASo は図 1.2 に示すように，製作するアプリケーションに応じて，レンダリングの種類を選択可能な構成をとっている．選択可能なレンダリング手法は以下の 4 種である．

1. 位置情報のみの高速レンダリング
2. 位置情報 + ドプラ効果の移動音像レンダリング
3. 位置情報 + 初期反射音のレンダリング
4. 位置情報 + 初期反射音 + 後部残響音のレンダリング

このように SiFASo は他の聴覚ディスプレイには見られない柔軟な構造をとっており，新たなレンダリングの実装を容易に行うことが可能という特長を持つ．以降本研究においては，SiFASo への実装を最終的な目的として検討を行うこととする．

1.2.3 回折音の表現

室内空間における音響効果の表現は，更なる忠実性を高めるための必須事項であると言える．また，その応用性も高く，将来的にはコンサートホールのシミュレーションによるバーチャルコンサート [19] や，視覚障害者のための空間認知訓練 [20] などのコンテンツへの利用が考えられる．

前述の SiFASo には，音源位置情報や反射音の付加機能が備わっているが，これらのシステムは自由空間，または直方体の室内空間に限定されたものであるため，廊下の曲がり角や，障害物のある空間など，直接音の届かない状況が発生する空間を対象とすることはできない．したがって，現状の SiFASo は十分に 3 次元音響特性を表現しているとはいえず，更なる忠実性向上のためにも，レンダリング可能な空間の拡張は必須であると考えられる．

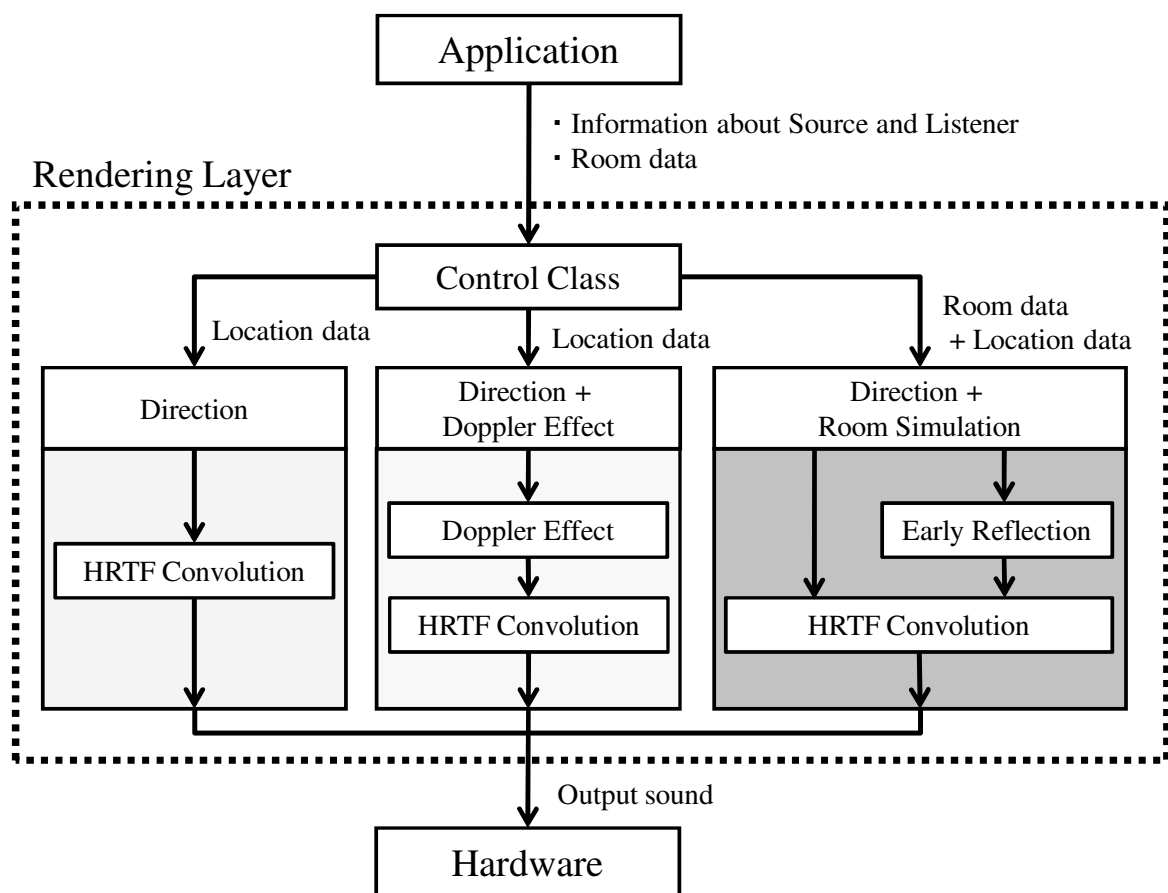


図 1.2: SiFASo で用意したレンダリング種類の構成

音源から放射された音は、聴取者との位置関係から伝達経路が決定し、最終的に聴取者の耳に到達する。ここで音源と聴取者の間に障害物が存在する場合、直接音が聴取者の耳に届くことはない。一方、現実の音空間を考えると、間に障害物がある場合でも、発せられた音は耳に届き実際に聞くことができる (図 1.3)。

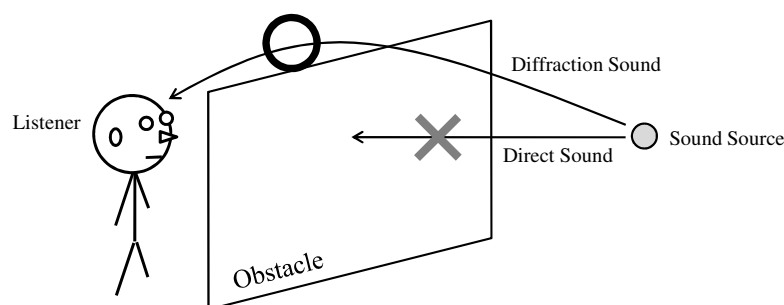


図 1.3: 音波の回折伝搬の概念図

これは、回折と呼ばれる現象であり、波がその障害物の背後など、つまり一見すると幾何学的には到達できない領域に回り込んで伝わっていく現象一般のことを言う。回折は音のみならず、光などの波動一般に対して発生する現象である。例えば、川岸から流れの中央に向かって流木があるような状況を考えると、川面に石を投げ入れたら波紋が生じるが、その波紋が流木を通過する時に、直接的には流木によって遮られているにも関わらず、流木の先端を回ってその反対側まで回り込んで伝わって行く。これと同様の現象が音に関して起こっているのである。

従来、回折音場は主に騒音対策の視点から解析、評価が行われてきており、回折音をいかに低減するかを主眼として研究が進められてきた [22, 23]。しかし近年、特にバーチャルリアリティの分野において、回折音は再現する対象としてとらえられることも多くなり、その手法に関する研究が注目を浴びてきている (例えば [24, 25])。

その一方で、従来の聴覚ディスプレイでは、SiFASo のように、聴取者または音源の移動に対応 (動的レンダリング) していても、回折音を実装したシステムは少なく、更なる高忠実レンダリングの実現のためには、回折音レンダリング手法の実装の必要性がある。

1.2.4 背景音の表現

既存の聴覚ディスプレイにおいては、通常、楽器の音や音声などの「聴取者に聞かせることを目的とした音 (以下、対象音と呼ぶ)」のみを出力している。一方で、我々が普段耳にする実際の音空間を考えると、これらの音のみが耳に入ってくるわけでは

なく、換気扇の駆動音や風の音などの「背景として存在はするが、聴取者が聞くことを目的としない音（以下、背景音と呼ぶ）」を含めた音空間を聴取している。

そのため、実際に対象音処理のみを行う聴覚ディスプレイによって提示される聴覚空間は、極めて静かな録音スタジオの中で、聴衆が自分しかいないコンサートを聴くかのような、聴感上非常に不自然なものとなる。したがって、提示される空間の忠実性を向上させるために、聴覚ディスプレイの出力として、通常の対象音に何らかの背景音を加えることが必要である。

しかし、現在までに、背景音に注意を払っているシステムは驚くほど少ない。数少ない例として、Lokki らは実験の際に使用する背景音に、ダミーヘッドを用いてバイノーラル録音した実際の背景音を使用している [21]。また、関らは背景音を対象範囲のはるか遠方に背景用の音源を配置し、反射などのレンダリング処理を施すことで、背景音が耳に入るようにレンダリングした視覚障害者向けの訓練システムを開発した [26]。しかし、実測した背景音を用いる場合、当然ながらその場所で実際に測定を行わなければならない、非常に煩雑になる上に、頭部運動に対応させるためには、それぞれの頭部の向きに対応した水平角、仰角に頭部を向けて測定をしなければならないため、背景音に時間的な不連続性が生じてしまう。加えて、そもそも実際には存在しないオリジナルな空間を作成する場合、実測はそもそも不可能である。背景音源を配置する場合においても、一般的に背景音源は無限に存在しているためその全てをレンダリングすることは不可能であり、また、有限個の背景音源を対象としてレンダリングした場合でも、それぞれに壁面反射などの処理を施すには多大な計算を必要とするとは明らかであり、そのために、対象音のレンダリングがおろそかになっては本末転倒である。これらの理由から、物理的に忠実な背景音のレンダリングは現状では困難であると考えられる。

そこで、本研究では背景音の検討に関して、物理的アプローチに加えて、知覚的アプローチを用いることとする。すなわち、物理現象として背景音をとらえるのではなく、背景音を付加することで私たちが普段実際に聞いている音空間に近い音空間を模擬し、知覚的に高忠実な背景音レンダリングを目指す。

1.3 研究の目的

以上に述べたような背景から、本研究では現状のシステムにおける音響効果の更なる高忠実化の実現を目的とし、聴覚ディスプレイの機能の新たなレンダリングアルゴリズムを、物理的アプローチ、知覚的アプローチの両面から提案する。また、急激な成長を遂げているコンテンツ産業のさらなる発展のために、3次元音響コンテンツの制作環境整備を目的としたミドルウェア SiFASo の機能拡張を行う。

まず第 2 章では，自由空間における回折音場の表現方法，及び先行研究をもとに最適なレンダリング手法の実装方法について検討を行ったのち，SiFASo のレンダリングエンジン部分に回折音処理スレッドクラスを追加する．

第 3 章では，主要な音の背景となる音に焦点をあて，背景音を効果的に表現するレンダリングアルゴリズムについて検討を行う．前述したように，既存の手法には柔軟性や演算量の面での多くの課題点が存在するため，これらとは異なる新たなアルゴリズムの提案が必要となる．本研究では背景音の付加が忠実性の感じ方にどのような影響を与えるかについて検証すると共に，計算負荷をできるだけ抑え，かつ効果的な背景音レンダリングアルゴリズムを提案する．

第 4 章では，聴取者の頭部運動が背景音を含む音空間の忠実性に与える影響について検討し，忠実性を高める背景音の条件について考察する．

最後に 3 次元音響技術及び本研究の展望について述べる

表 1.1: 代表的な聴覚ディスプレイとその特徴

システム開発者 (システム名)	言語	OS	遅延 ¹	更新レート ²	付加効果
Wenzel <i>et al.</i> (SLAB)[11]	C++	Windows	25 ms	120 Hz	音源位置-聴取者 3 次元位置， 直方体室 1 次反射 (虚像法)， 空気吸収，壁面素材 (吸収)
Savioja <i>et al.</i> (DIVA)[10]	C++	UNIX， Linux	110-160 ms	20 Hz	音源位置-聴取者 3 次元位置， 不整形室 2 次反射 (虚像法)，残響，空気吸収， 壁面素材 (吸収)，距離減衰，ドップラー効果
OpenAL[12]	C++ 他	Windows 他	150 ms	機種依存	音源-聴取者 3 次元位置， 残響，空気吸収，距離減衰，ドップラー効果
Yairi <i>et al.</i> [13]	C	Linux	12 ms	120 Hz	音源-聴取者 3 次元位置 (両者を能動的に指定可能)， 距離減衰
Toyoda <i>et al.</i> (SiFASo)[9]	C++	Windows	40 ms	60 Hz	音源-聴取者 3 次元位置， 直方体室 1 次反射 (虚像法)，残響， 距離減衰，空気吸収，壁面素材 (吸収)， ドップラー効果 (位置情報付加で自動的に)

¹ 聴取者の頭部運動などの変化が反映された音出力されるまでの遅延の合計

² 使用するセンサによる位置情報の更新レート．センサ性能及びアルゴリズムに依存

第2章 回折音のレンダリングに関する検討

2.1 はじめに

SiFASo には、頭部伝達関数合成法による音源位置情報の表現や、室内における反射音の付加を行うための機能が備わっているが、現状において、これらのシステムは自由空間、または直方体の室内空間に限定されたものであるため、廊下の曲がり角や、障害物のある空間など、直接音の届かない状況が発生する空間を対象とすることはできない。したがって、現状の SiFASo は十分に 3 次元音響特性を表現しているとは言い難く、更なる忠実性向上のためにも、レンダリング可能な空間の拡張は必須であると考えられる。

本章ではまず過去に提案された回折音のレンダリング手法について概観し、検討する方法を考察する。本研究の回折音のレンダリングは、まず可能な限り物理的に忠実であることを目指す。そこで、本研究では回折現象の最も基本的な理論である、Huygens の原理をもとに考案された計算法に限定して検討を行う。また、最終的に実装処理を施すことを考えると、アルゴリズムが高速で、かつレンダリング可能な環境を柔軟に設定できることが望まれる。以上の条件を満たす、最適な手法を考察する。

続いて、選定された手法を SiFASo のレンダリングクラスに追加する。実際にミドルウェア構築を行うことを考えると、使用するレンダリングアルゴリズムに対して、模擬精度と計算負荷の両者を考慮に入れた設計が必要となる。本章では、現在の一般的な計算機の処理能力を考慮に入れた上での、具体的なアルゴリズムの設計法について述べる。

最後に、SiFASo に組み込んだ回折音レンダリングアルゴリズムの模擬精度に関する評価を行う。無響室において回折効果を含む音をダミーヘッドを用いて実際に録音し、SiFASo の出力音との比較を行うことで精度を評価する。

2.2 Huygens の原理

音が障害物の陰の部分にも伝搬していく回折現象は、Huygens の原理 [7] によって説明される。Huygens の原理とは、波動の同じ位相の点を連ねた面として定義され

る波面の伝搬を説明するもので、それによると、波面が無数個の微小音源で構成されるとし、ある時刻における波面上の各点から放射された 2 次波が重なり合った結果として、次の波面が作られる。これは Huygens が光の波動説において、反射・屈折の原理を説明するために考えたもので、後に Fresnel が回折現象をこれによって説明し、さらに Kirchhoff が波動方程式をもとにして数学的に証明した。この理論は現在に至るまで、波動における基本的な定理の一つであり、回折音のレンダリング手法においても、最も基本的な考え方となっている。

平面波は、微小点波源が平面上に密接に並んでいるものと考えることができる。その波面が、図 2.1(a) の左側の直線上にあるとすると、各微小点波源からの波面は、時間の経過により各点波源を中心とする球面に広がるが、包絡線はやはり平面で、波の進行方向は変わらない。

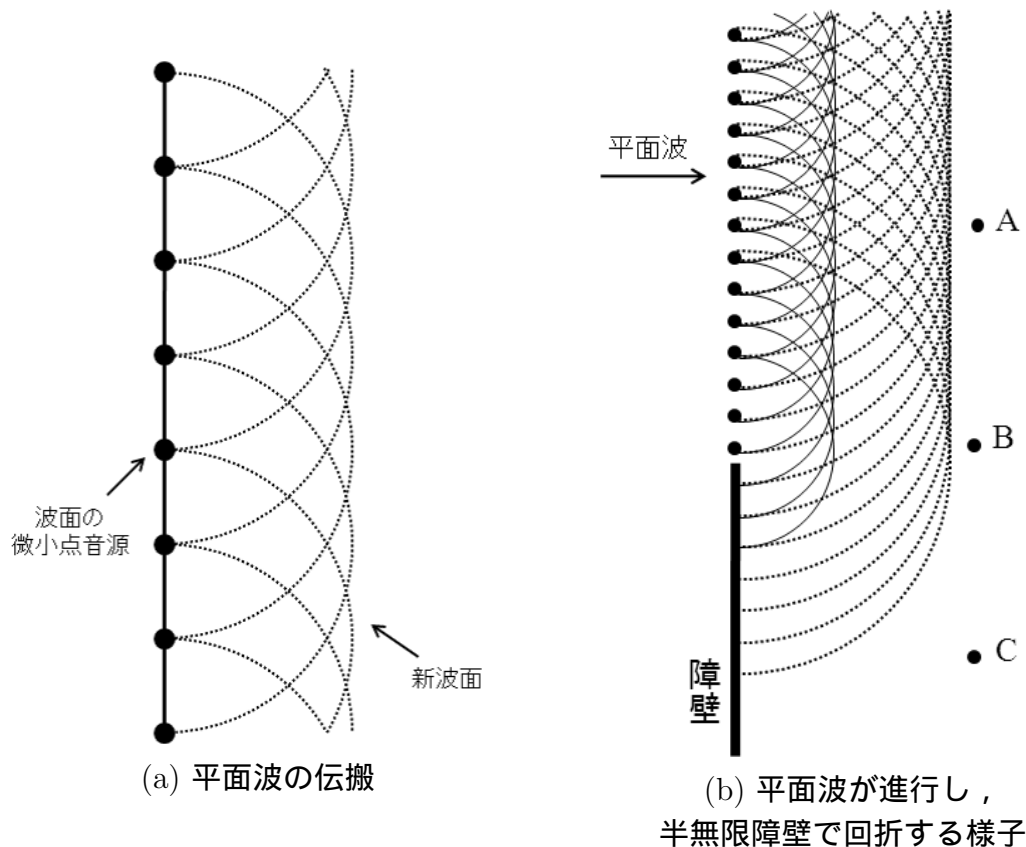


図 2.1: Fresnel-Huygens の原理

球面波の場合も同様に考えることができる。このときには、一定の時間経過した時の波面は、前の波面と同じ中心の球面である。図 2.1(b) を用い、平面波の進行方向に垂直な半無限障壁の裏側に、音波が回折していく現象を考える。壁面の延長面の、ある時点の波面に並ぶ各微小点音源からの音波の音圧は、音源からの距離に反比例し、距離相当分の位相の遅れを持って広がる。

ある時間 τ 後の波面は，音速を c とすると，元の波面の各点を中心にした半径 $c\tau$ の球面を連ねた面になるが，その面上の各点の音圧は，そこに到達した音圧の和になる．

したがって，障壁から遠い点 A 付近では，平面波に近く，伝搬によって音圧はあまり下がらないが，障壁に近い点 B では，位相差が大きい音波しか重なり合わなくなるために，音圧が小さくなり，陰になる奥の点 C ではそれが著しくなる．

波長の長い低周波数の音波は，距離による位相の回転が少ないため，障壁の陰においても音圧の低下が少ないが，距離による位相回転が大きい高周波数の音波は，障壁の陰に入ると音圧が急激に低下する．ここでは無視している障壁からの反射音も，このような音圧の違いの原因の一つになる．

2.3 回折音場のレンダリングに関する検討

2.3.1 過去に提案された回折音場のレンダリング手法

波長に比べ十分大きな孔などによる球面波の回折現象を説明した古典的理論としては，Huygens の原理を応用した Fresnel の輪帯理論，並びに Huygens の原理を数学的に厳密な積分公式とした上で求められた Kirchhoff の近似解が知られ，現在までよく使われている [27]．一方，それらの近似回折理論に対し，厚みのない半無限障壁に平面波が入射した場合の厳密な境界条件に基づく厳密解が，Sommerfeld によって求められた [28]．また，球面波入射に対しても，Sommerfeld の方法を基にして Carslaw が解法を示した [29]．更に Macdonald は厳密解の整理された形式を与えた [30]．ところが，遮音壁の設計のように，障害物による減衰量計算には厳密解がほとんど使われておらず，前川の実験曲線 (前川チャート) [31] が広く使用されてきた．しかしその実験曲線は，Kirchhoff の近似理論の適用範囲内で行われた実験に基づいて Kirchhoff の近似値を補正する形で求められたものであり，適用範囲も限定されている．それらの適用範囲外である障壁上に音源が分布する場合も，音波の伝搬を考える際に非常に重要な基本モデルであり，全領域への音波の伝搬を取り扱う必要がある．

このような観点から，近年，音源及び受音点を問わず適用可能な手法が考案されている．以下では，主な回折音のレンダリング手法の概要について述べる．

堀内らの手法

堀内，東山らは，波長に比べ厚みが無視できる半無限障壁に対して，第一種 Hankel 関数を用いて回折音場を表現している [32]．この手法は，前述の Macdonald の解法

に着目したものであり，半無限障壁に限定された手法ではあるものの，可能な限り解析的に厳密な解を目指した解法である．

図 2.2 に示すように，波長に比べ厚みが無視できる半無限障壁 ($0 < x, y = 0$) を考える．

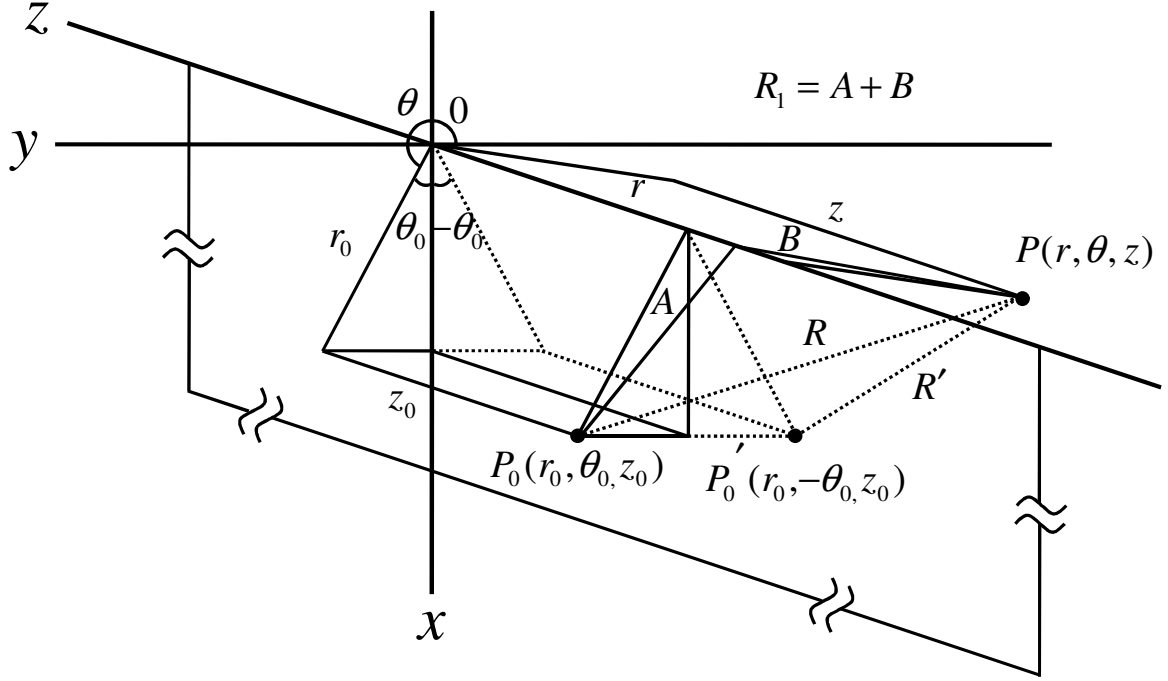


図 2.2: 堀内らの回折シミュレーション環境

円筒座標系で点 $P_0(r_0, \theta_0, z_0)$ に点音源があるとき，音源から出る球面波による点 $P(r, \theta, z)$ の速度ポテンシャル $\Phi = \phi e^{j\omega t}$ は，時間項を除いた入射波面波を $\phi_i = \frac{e^{-jkR}}{R}$ とすると

$$\phi = U(\theta_0) + U(-\theta_0) \quad (2.1)$$

と表わされる．ここで， $U(\theta_0)$ ， $U(-\theta_0)$ はそれぞれ音源と鏡像による受信点の速度ポテンシャルを示す．また U は幾何音響学的音波 U_g を回折波 U_d で，それぞれ

$$U_g(\theta_0) = \begin{cases} \frac{e^{-jkR}}{R} & \text{for } \cos \frac{1}{2}(\theta - \theta_0) \geq 0 \\ 0 & \text{for } \cos \frac{1}{2}(\theta - \theta_0) < 0 \end{cases} \quad (2.2)$$

$$U_d(\theta_0) = \pm jk \int_{|\tau_R|}^{\infty} \frac{H_1^{(1)}(\tau^2 + kR)}{\sqrt{\tau^2 + 2kR}} d\tau \quad \pm \text{ for } \cos \frac{1}{2}(\theta - \theta_0) \gtrless 0 \quad (\text{複合同順}) \quad (2.3)$$

と表わされる．ただし，

$$\begin{aligned}\tau_R &= \pm \sqrt{kR_1 - kR} \quad \pm \text{ for } \cos \frac{1}{2}(\theta - \theta_0) \gtrless 0 \\ \tau'_R &= \pm \sqrt{kR_1 - kR'} \quad \pm \text{ for } \cos \frac{1}{2}(\theta + \theta_0) \gtrless 0\end{aligned}$$

$$\begin{aligned}R &= \sqrt{r^2 + r_0^2 + (z - z_0)^2 - 2rr_0 \cos(\theta - \theta_0)} \\ R' &= \sqrt{r^2 + r_0^2 + (z - z_0)^2 - 2rr_0 \cos(\theta + \theta_0)} \\ R_1 &= \sqrt{(r + r_0)^2 + (z - z_0)^2}\end{aligned}$$

$$k = \frac{\omega}{c} : \text{波定数}$$

である．ここで $H_1^{(1)}$ は第 1 種 Hankel 関数， R 及び R' はそれぞれ実音源 P_0 及び虚音源 P'_0 点から観測点 P までの距離， R_1 は P_0 からエッジを乗り越えて P 点へ到達する最短距離である．

$1 \ll kR_1$ のとき，式 (2.3) は Hankel 関数をその漸近展開表示式の第一項で近似することで，

$$\begin{aligned}U_d(\theta_0) &\simeq \pm jk \sqrt{\frac{2}{\pi}} e^{-jkR + j\frac{\pi}{4}} \int_{|\tau_R|}^{\infty} \frac{e^{j\tau^2}}{\sqrt{(\tau^2 + kR)(\tau^2 + 2kR)}} d\tau \\ &\quad \pm \text{ for } \cos \frac{1}{2}(\theta - \theta_0) \gtrless 0\end{aligned} \quad (2.4)$$

と表すことができる． $\pi \ll kR_1$ のとき，式 (2.4) の積分記号内の $\sqrt{(\tau^2 + kR)(\tau^2 + 2kR)}$ は $e^{j\tau^2}$ に比べ非常にゆっくり変化する項であり， $\sqrt{(\tau^2 + kR)(\tau^2 + 2kR)}$ として積分の外に出しても十分な精度を得られる [22]．したがって複素 Fresnel 積分 [33]

$$F(a) = \int_a^{\infty} e^{j\tau^2} d\tau \quad (2.5)$$

を用いて

$$\begin{aligned}U_d(\theta_0) &\simeq U_g(\theta_0) \pm j \sqrt{\frac{2}{\pi R_1(R_1 + R)}} e^{-jkR + j\frac{\pi}{4}} F(|\tau_R|) \\ &\quad \pm \text{ for } \cos \frac{1}{2}(\theta - \theta_0) \gtrless 0\end{aligned} \quad (2.6)$$

と近似される．音圧 $p(k)$ は

$$p(k) = \rho \frac{\partial \Phi}{\partial t} \quad (2.7)$$

$$= \rho j \omega \phi e^{j\omega t} \quad (2.8)$$

と表わされる．ここで ρ は気体密度， ω は角周波数である．

次に，サンプリング定理を満たすように帯域制限した $p(k)$ を $H(k)$ と置きなおす． $H(k)$ の実部 $H_r(k)$ を偶関数 $H_{reven}(k)$ ，虚部 $H_i(k)$ を奇関数 $H_{iodd}(k)$ として

$$H(k) = H_{reven}(k) + iH_{iodd}(k) \quad (2.9)$$

と表わし，回折波を考慮したインパルス応答 $h[n]$

$$h[n] = \frac{1}{N} \sum_{k=0}^{N-1} e^{j(\frac{2\pi}{N})kn} \quad (2.10)$$

を得る．

Medwin らの手法

Medwin らは BTM 法 (Biot-Tolstoy-Medwin solution) という，Biot-Tolstoy のインパルス応答算出法 [34] に基づいた手法を考案した [35]．インパルス応答は式 (2.11) で表わされる．この手法は計算機上での実装を意識して設計された手法であり，回折音の再現性と演算の高速化を兼ね備えた手法であると言える．

$$h_{BTM}(t) = -\frac{c\nu}{2\pi r_S r_R \sinh(\eta_{BTM})} \beta_{BTM} H(t - t_0) \quad (2.11)$$

ここで，

$$\beta_{BTM}(t) = \beta_{BTM++} + \beta_{BTM+-} + \beta_{BTM-+} + \beta_{BTM--}$$

$$\beta_{BTM\pm\pm}(t) = \frac{\sin[\nu(\pi \pm \theta_s \pm \theta_R)]}{\cosh(\nu\eta_{BTM}) - \cos[\nu(\pi \pm \theta_s \pm \theta_R)]} \quad (\text{複合同順})$$

$$\eta_{BTM}(t) = \cosh^{-1} \frac{c^2 t^2 - (r_S^2 + r_R^2 + (z_R - z_S)^2)}{2r_S r_R}$$

$$H(t) = \int_{-\infty}^t \delta(\tau) d\tau$$

である．各パラメータは図 2.3 に示す． c は音速， ν は障壁の角度を表す値 (Wedge Index) で， $\nu = \frac{\pi}{\theta_w}$ で与えられる． t_0 は音源から発せられた音が最短経路で受音点まで達するまでの時間を表す．また， $H(t)$ は Heviside の階段関数である．

計算機上での表現のためには， $h(t)$ を離散化した形で表現する必要がある．一般的に離散化されたインパルス応答 $h[n]$ は， $h(t)$ の $(n - 0.5)/f_s$ から $(n + 0.5)/f_s$ までの積分を用いて表わされる．

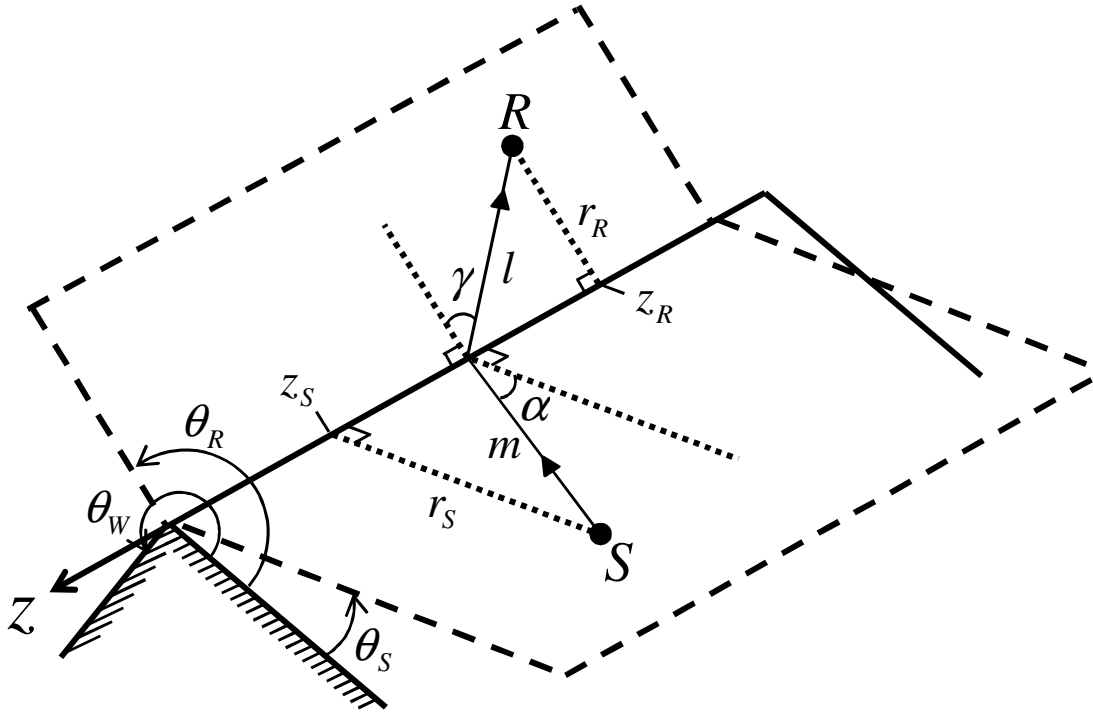


図 2.3: 1 次回折シミュレーション環境

$$h_{BTM}[n] = \int_{(n+0.5)/f_s}^{(n+0.5)/f_s} h(t) dt \quad (2.12)$$

Calamia-Svensson らの手法

Calamia らは高速かつ適用可能な環境が柔軟な回折音レンダリング手法を考案した [41]。回折音に関する高速化手法に関しては，Funkhouser ら [39] や Tsingos [40] らによって検討が行われてきたが，これらの高速化手法は，無限長の障壁にのみ対応しており，また音源，聴取者位置が十分に遠いことが前提条件として必要になってくる．これに対し，Calamia の手法 [41] は有限長の障壁に対応しており，音源，聴取者位置にも制限がない．

また，回折音の解析的算出法については，Svensson らの手法を用いている [36]．Svensson らの手法は，前述の Medwin らの手法 [35] が解析領域の境界付近で解に特異性を持つことを示した上で，この特異性を解消する ES(Edge Source) 法を提案するものであり，また，Svensson らは，ES 法が有限長障壁や多段回折に拡張可能であることを示している [38]．

ES 法を用いたインパルス応答は式 (2.13) で表わされる．

表 2.1: 各レンダリング手法のまとめ

	有限長障壁	多段回折	特異点の有無	演算速度
堀内らの手法	対応なし	対応なし	特異点なし	120 ms/position
Medwin らの手法	対応あり	対応あり	特異点あり	14 ms/position
Calamia-Svensson の手法	対応あり	対応あり	特異点なし	5.5 ms/position

$$h_{ES}(t) = -\frac{\nu}{4\pi} \int \delta\left(t - \frac{m+l}{c}\right) \frac{\beta_{ES}}{ml} dz \quad (2.13)$$

ここで, $\delta(t)$ は Dirac のデルタ関数であり,

$$\begin{aligned} \beta_{ES}(t) &= \beta_{ES++} + \beta_{ES+-} + \beta_{ES-+} + \beta_{ES--} \\ \beta_{ES\pm\pm}(t) &= \frac{\sin[\nu(\pi \pm \theta_s \pm \theta_R)]}{\cosh(\nu\eta_{ES}) - \cos[\nu(\pi \pm \theta_s \pm \theta_R)]} \quad (\text{複合同順}) \\ \eta_{ES}(t) &= \cosh^{-1} \frac{1 + \sin \alpha \sin \gamma}{\cos \alpha \cos \gamma} \end{aligned} \quad (2.14)$$

である.

インパルス応答の離散化手法は一般的に式 (2.12) が用いられるが, ES 法では次式のようにこれを dt に対応するエッジ dz 上の積分に変換して処理する.

$$h_{ES}[n] = \int_{z_{n2}}^{z_{n1}} h(z) \frac{dt}{dz} dz \quad (2.15)$$

2.3.2 本研究で採用する回折音レンダリング手法の詳細

以上の検討をまとめた結果を, 表 2.1 に示す.

本研究では, レンダリング可能な環境の種類やリアルタイム性などから, Calamia-Svensson らの手法を採用することとする. 以下に, 式 (2.13) で与えられるインパルス応答の離散化手法についての要点を述べる.

一般的に, 時間的連続関数であるインパルス応答 $h(\tau)$ は, $\tau = n/f_s$ を用いて離散化することによって, 式 (2.16) のように表わされる. ここで f_s はサンプリング周波数, n はサンプリングナンバーである.

$$h[n] = \int_{(n+0.5)/f_s}^{(n+0.5)/f_s} h(\tau) d\tau \quad (2.16)$$

式 (2.16) において、 $h(\tau)$ は ES 法における連続時間の回折音インパルス応答を示す。式 (2.16) においては、 $\tau = \tau_0$ においての特殊解を持つことに注意しなければならないため、 τ_0 における値は、 τ_0 の近傍の τ の値によって近似される。

回折音を、エッジに複数配置した微小音源から発せられる音波の和として表現する場合、微小時間 $d\tau$ 上の積分である式 (2.16) は、 $d\tau$ に対応したエッジの微小区間 dz 上の積分に等しいと言える。式 (2.16) を dz 上の積分に変換し、式 (2.17) を得る。

$$h[n] = \int_{(n+0.5)/f_s}^{(n+0.5)/f_s} h(\tau) d\tau = \int_{z_{n2}}^{z_{n1}} h(z) \frac{d\tau}{dz} dz \quad (2.17)$$

z_{n1} 及び z_{n2} は、それぞれ $\tau = (n+1)/2f_s$ 、 $\tau = (n-1)/2f_s$ に対応する z 軸上の点である。これにより、式 (2.14) から以下の関係が得られる。

$$h(z) \frac{d\tau}{dz} dz = -\frac{\nu}{4\pi} \frac{\beta[\alpha(z), \gamma(z), \theta_R, \theta_R]}{m(z)l(z)} dz \quad (2.18)$$

式 (2.18) を用いることで、式 (2.17) は、式 (2.19) のように表わすことができる。

$$h[n] = -\frac{\nu}{4\pi} \int_{(n+0.5)/f_s}^{(n+0.5)/f_s} \frac{\beta[\alpha(z), \gamma(z), \theta_R, \theta_R]}{m(z)l(z)} dz \quad (2.19)$$

ここで、式 (2.19) は中間点の値を用いることで式 (2.19) のように近似することができる。

$$h[n] \approx -\frac{\nu}{4\pi} \frac{\beta[\alpha(z_n), \gamma(z_n), \theta_R, \theta_R]}{m(z_n)l(z_n)} \Delta z_n \quad (2.20)$$

z_n は z_{n1} と z_{n2} の中間点である。ここで、演算の簡略化のため、エッジを等間隔 Δz に分割すると、要素 i によって、 z_i に対応するインパルス応答 Δh_i は次式で与えられる。

$$\Delta h_i \approx -\frac{\nu}{4\pi} \frac{\beta[\alpha(z_i), \gamma(z_i), \theta_R, \theta_R]}{m(z_i)l(z_i)} \Delta z \quad (2.21)$$

エッジの分割は各エッジ要素の位置と区間長によって決定される。 Δh_i による寄与は、要素長 Δz と、位置 z_i によって与えられた、時間幅 Δt の方形パルスと見ることができる。

表 2.2: SiFASo の実装環境

Operating System	Microsoft Windows XP Professional Edition
CPU	Pentium 4 Extreme Edition 3.2 GHz
Memory size	2 GB
Audio interface	DirectX(ver 9.0), ASIO
Sound Card	M-Audio AudioPHILE 2496
Compiler Language	Microsoft VisualStudio .Net 2003
Programing	C++

2.4 SiFASo への実装

豊田らは表 2.2 に示される環境において SiFASo を実装している．そこで，本研究においても同環境において稼働可能となるようにレンダリングアルゴリズムの実装を目指す．

SiFASo のレンダリングエンジン部分のクラス構成について図 2.4 に示す (各クラスの詳細な役割は [9] 参照)．本研究では従来までのクラスに加えて，以下の二つのクラスを追加する．図 2.5 に実装する回折音レンダリングアルゴリズムの一連の処理過程を示す．まず，入力情報として，聴取者情報 (位置 $P_R(x_R, y_R, z_R)$ ，頭部方向 $A_R(yaw, pitch, roll)$)，音源情報 (位置 $P_S(x_S, y_S, z_S)$ ，音源信号 $S(t)$)，障壁情報 (位置 $P_W(x_W, y_W, z_W)$ ，角度 θ_W) が与えられる．SiFASo のデフォルトの座標系はデカルト座標なので，Calamia-Svensson の手法を用いるために円柱座標に変換したのち，回折音場のインパルス応答を計算する．次に各位置情報，頭部方向から適切な HRIR を選択する．ここで，適切な HRIR が存在しない場合は，最近接の 4 点の HRIR から，線形補間により算出する．得られた HRIR と，前述の回折音インパルス応答を畳み込むことで，合成インパルス応答を算出し，さらにこの合成インパルス応答と音源信号を畳み込むことで，方向特性を付加した回折音を得る．

- CVADDifThread(回折音付加スレッドクラス)

CVADPCMThread の派生クラス．直接音の表現に加え，障壁による回折音を模擬する．直接音と回折音のミキシングは本クラス内で行われ，外のクラスに影響を及ぼさない．

- CVADDif(回折音処理クラス)

回折音付加に関するすべての具体的な音響処理を行うクラス．

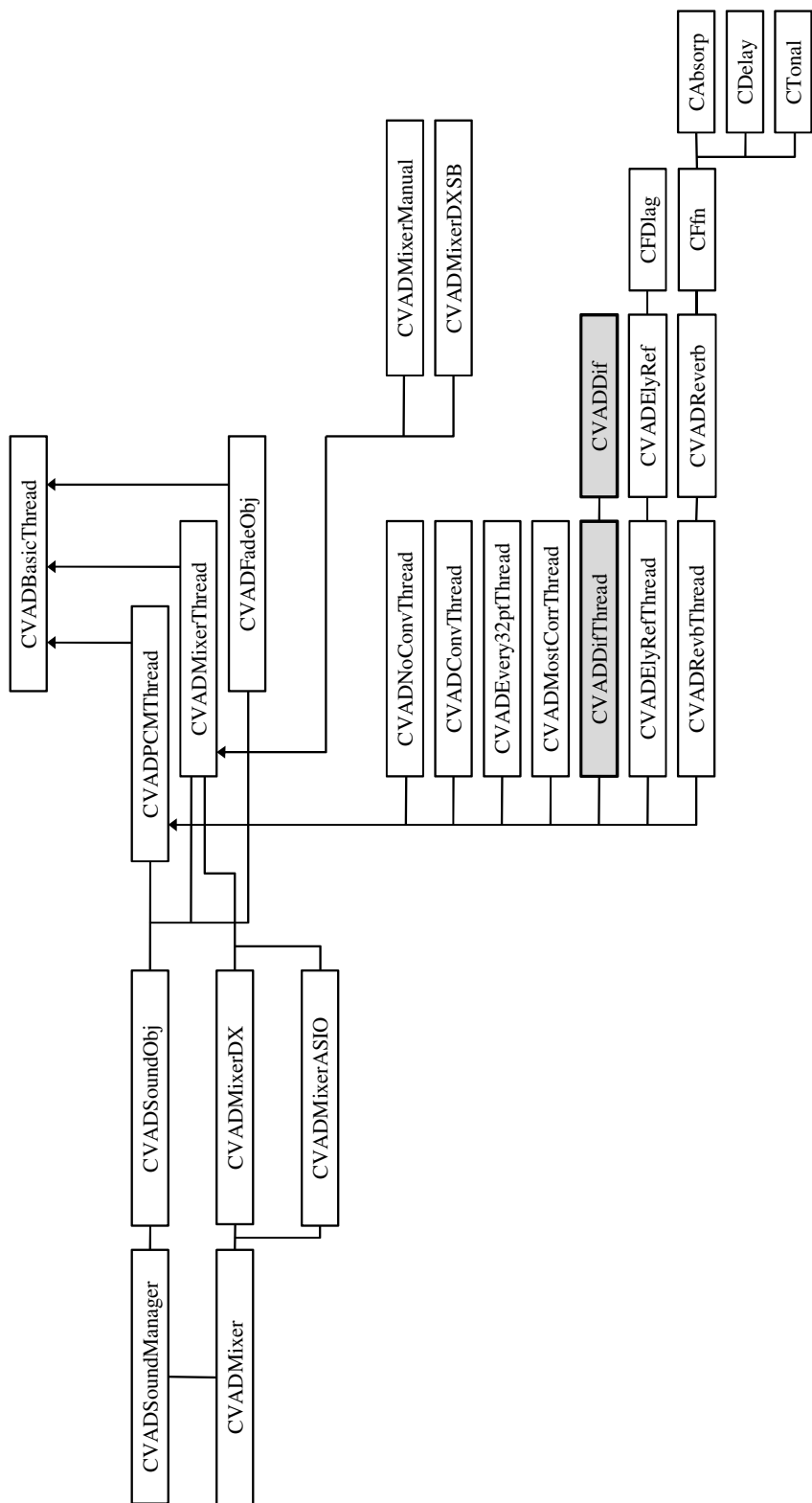


図 2.4: SiFASo のレンダリング部分のクラス構造

表 2.3: ラウドスピーカ特性の測定条件及び測定系

測定信号	OATSP[37]
サンプリング周波数	48 kHz
サンプル長	8192 pt
折り返し歪み除去時の遮断周波数	22 kHz
同期加算回数	10
Microphone	Knowles FG3329
Microhone Amplifer	Thinknet MA-2016
D/A A/D Converter	PAVEC MD-8000 mk2
Speaker Amplifer	ONKYO TOKKI MCA8050

2.5 レンダリングアルゴリズムの評価

本検討において実装すれたレンダリングアルゴリズムがどの程度の精度を有しているかについての検討を行う．そこで，実際に回折音場の測定を行い，計算機シミュレーションによって求められた結果との比較を行う．

2.5.1 測定に使用するラウドスピーカについて

回折音場の測定を行う際，壁を回り込んで音が伝わることを考えると，ラウドスピーカは可能な限り無指向性であることが望まれる．そこで，本測定では多面体ラウドスピーカを使用することとした．

はじめに，実験で使用する多面体ラウドスピーカを選定するため，Solid Acoustics の 12 面体ラウドスピーカ（図 2.6(a)）と小貫ら [42] らによって作成された 32 面体ラウドスピーカ（図 2.6(b)）の 2 つのラウドスピーカの性能を比較する．

測定は電気通信研究所無響室内で行う．測定条件，測定系を表 2.3 に示す．

各多面体ラウドスピーカの周波数特性を図 2.7 に示す．なお，この特性にはマイクロホン (Knowles FG3329) の周波数特性も含まれており，それぞれ $f = 100\text{Hz}$ の値を 0 dB として正規化している．

測定データを比較した結果，低域の安定性の面で 12 面体スピーカが優れていた．32 面体スピーカには，点音源に近いというがあるが，出力できる音のレベルが比較的小さく，本実験のような環境における測定には不向きであると言える．そこで，本実験では以降 Solid Acoustics の 12 面体スピーカを使用する．

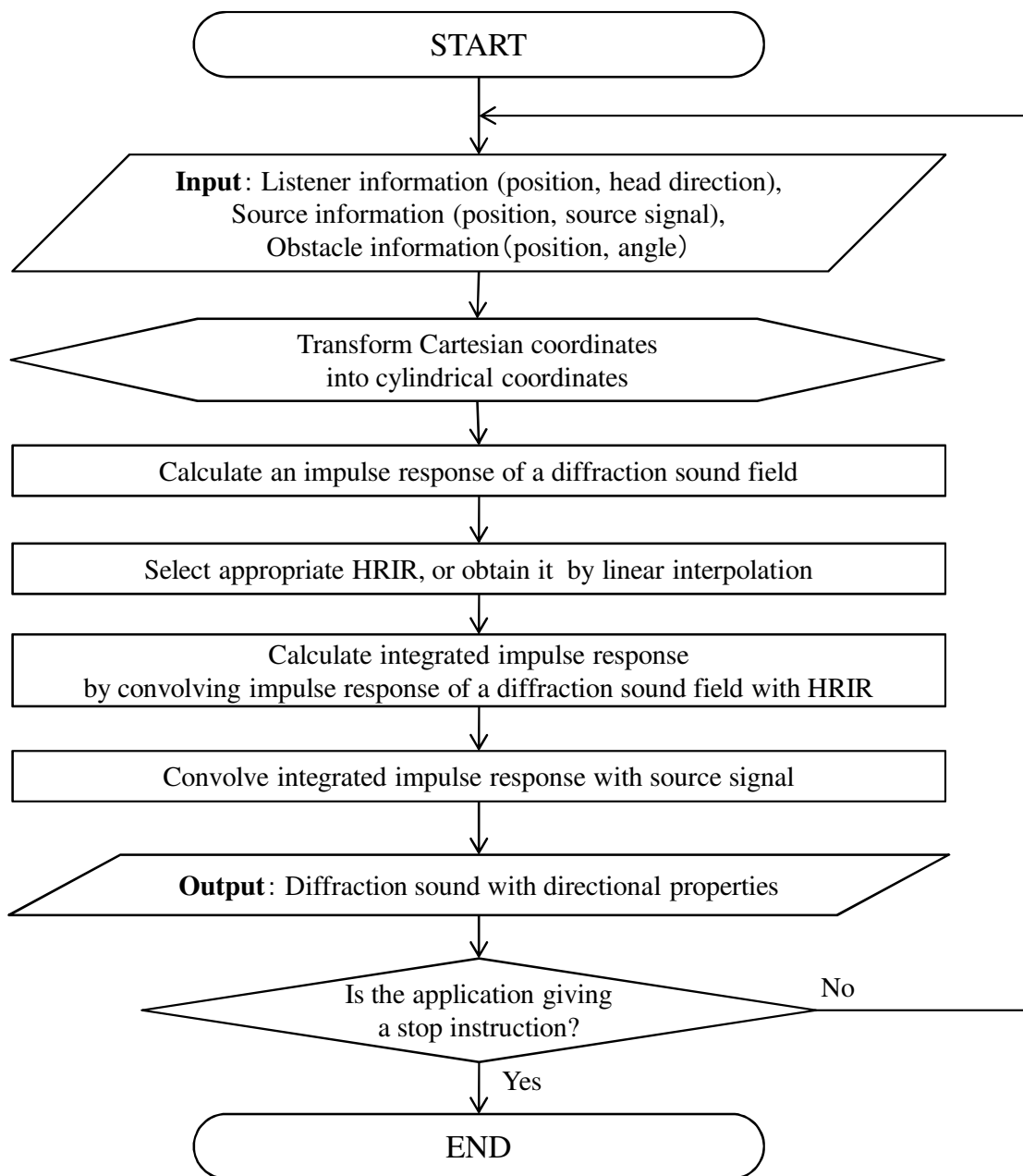


図 2.5: 回折音レンダリングの処理フローチャート

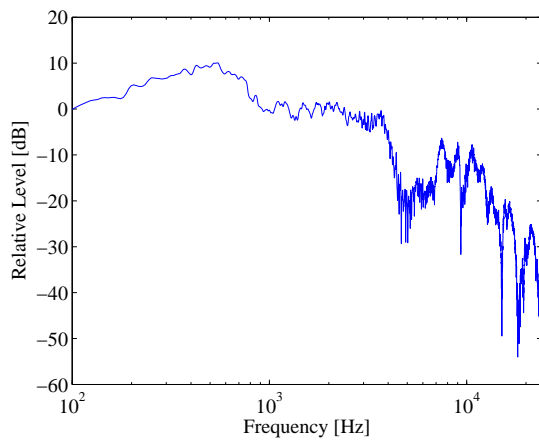


(a) 12 面体ラウドスピーカ

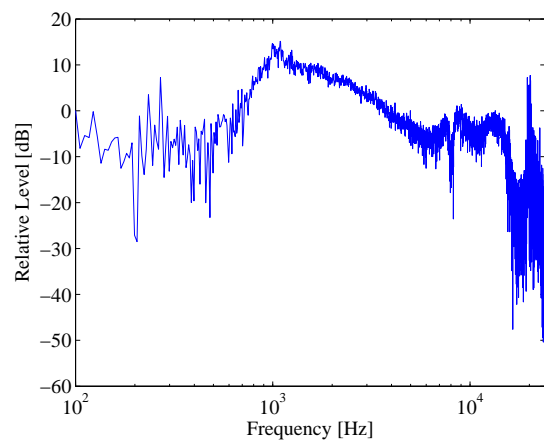


(b) 32 面体ラウドスピーカ

図 2.6: 測定に用いた多面体ラウドスピーカ



(a) 12 面体ラウドスピーカ (Solid Acoustics) の周波数特性



(b) 32 面体ラウドスピーカ (小貫ら) の周波数特性

図 2.7: 各多面体ラウドスピーカの周波数特性

2.5.2 回折音場の測定とシミュレーション結果の比較

一例として，図 2.8 のシミュレーション環境について，実際に周波数特性を求めた．ここでは比較のため，無響室において実際に木材の板を障壁として回折音を測定した際の環境を再現している．測定条件，測定系は表 2.3 に示す条件と同様の条件で測定を行った．

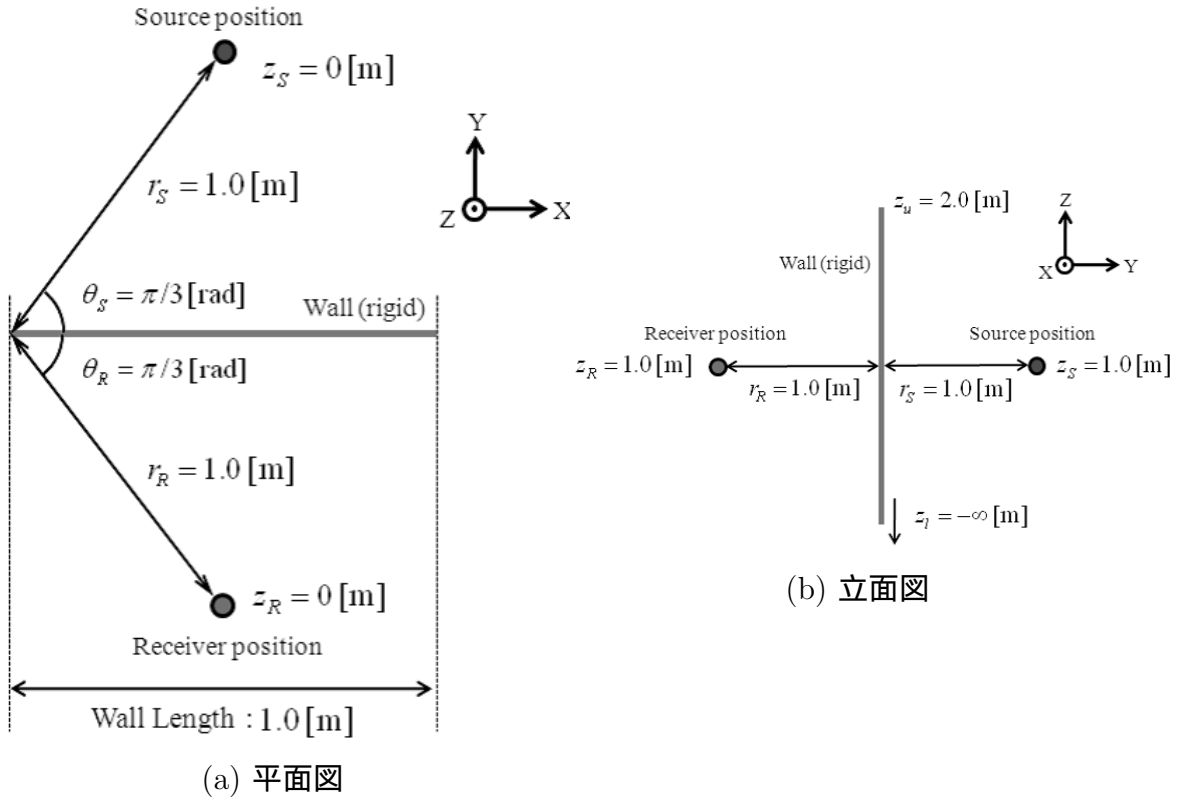


図 2.8: シミュレーション環境

Calamia-Svensson の手法による計算機シミュレーションによるデータ (図 2.9(a)) と実測データ (図 2.9(b)) の周波数特性を以下に示す．それぞれのデータはインパルス応答の開始点から 1024 ポイント (21.3 ms) の時間フレームで切り出したものを用いている．また両図は，壁を取り除いた状態との音圧の比を表わしている．すなわち，各周波数における基準音圧 (0 dB) は，壁が存在しない場合の直接音の音圧である．

実測と計算機シミュレーションとの周波数特性を比較すると，低域部においては両者の特性は似通っているものの，計算機シミュレーションの結果は実測の結果における高域部の減衰を表現しきれていないことが見て取れる．

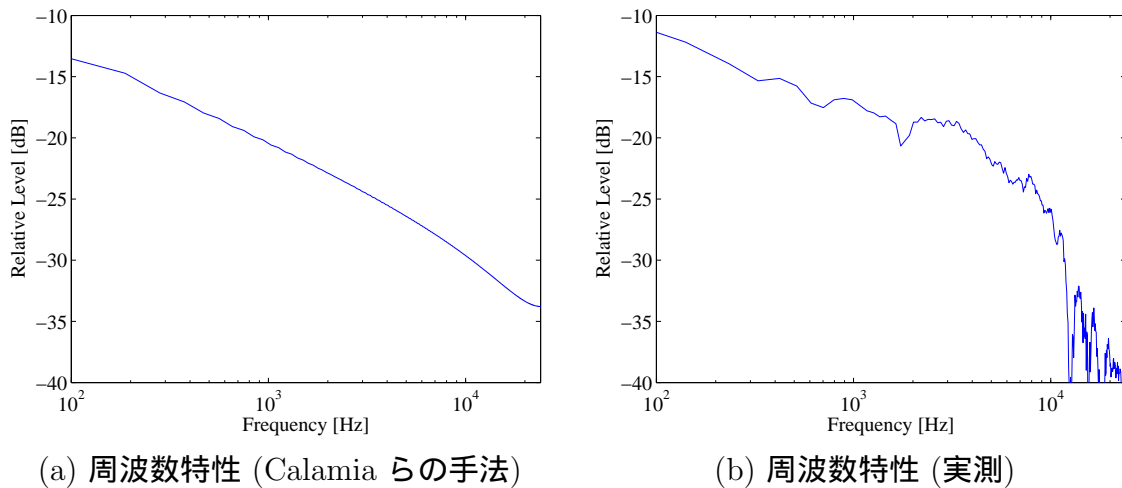


図 2.9: 周波数特性による Calamia-Svensson の手法の評価

2.6 まとめ

現在の SiFASo には、頭部伝達関数合成法による音源位置情報の表現や、室内における反射音の付加を行うための機能が備わっているが、これらのシステムは自由空間、または直方体の室内空間に限定されたものであるため、廊下の曲がり角や、障害物のある空間など、直接音の届かない状況が発生する空間を対象とすることはできず、更なる忠実性向上のためにレンダリング可能な空間の拡張は必須であると考えられた。

本章では、このような空間をレンダリング可能とするため、回折音場の最適レンダリング手法について検討を行った。回折音のレンダリング手法には、演算結果が実空間に忠実であることに加え、演算可能な環境が柔軟であること、及び演算が高速であることが求められる。これらの条件から様々な回折音レンダリング手法について検討を行った結果、本研究では Calamia-Svensson の手法を採用することとした。Calamia-Svensson の手法は、高速化アルゴリズムの適用時に音源と聴取者の位置に制限がないため、柔軟性に優れ、演算も非常に高速であった。また、Calamia らが回折音の演算部分に使用している Svensson らの手法は、有限長の障壁や、多段回折音に対応可能であり、拡張性に優れている。

続いて Calamia-Svensson の手法を豊田らの開発した SiFASo に実装した。SiFASo に実装されている他のレンダリングアルゴリズムに習い、パラメータとして聴取者の位置情報及び頭部の方向情報、音源の位置情報、及び障壁の配置位置情報を入力することで、方向特性を付加した回折音を出力する仕様とした。それに伴い、SiFASo のレンダリングエンジン部分に回折音付加スレッドクラスである、CVADDifThread、回折音処理クラスである、CVADDif を追加した。

本手法を無響室にて測定した実測の回折音と比較した結果、低域部の減衰は比較的

よく表現されているものの，Calamia-Svensson の手法のみでは高域部の減衰を表現しきれていないことがわかった．

第3章 背景音のレンダリングに関する検討

3.1 はじめに

対象音の背景として存在する音を提示する際に実測した背景音を用いる場合，実際にその空間で測定を行わなければならない，工程が非常に煩雑になる．加えて，頭部運動に対応させるためには，それぞれの頭部の向きに対応した水平角，仰角に頭部を向けて測定をしなければならないため，背景音に時間的な不連続性が生じてしまう．また，実際には存在しないオリジナルな空間を作成する場合，実測自体が不可能となる．背景音源を対象音と同様に，擬似的に配置する場合においても，一般的に背景音源は無限に存在しているためそのすべてをレンダリングすることは不可能であり，有限個の背景音源を対象としてレンダリングした場合でも，それぞれに壁面反射などの処理を施すには多大な計算を必要とすることは明らかである．これらの理由から，物理的に忠実な背景音のレンダリングは現状では困難であると考えられる．そこで本章では，背景音のレンダリングを行うにあたって，物理的アプローチではなく，知覚的アプローチによる新しいアルゴリズムを設計する．

まず，実空間における背景音を基に，背景音源として最適な音源について検討する．この背景音を用いて，背景音の付加手法についての検討を行う．また，背景音を付加することが聴取者が感じる音空間全体の忠実性にどのような影響を及ぼすかを検証する．

3.2 背景音に関する先行研究

吉田ら [43] は立体音の本質は空間構造の再現にすべてが帰せられ，この空間構造は，対象音を空間的配置から選択聴取する方向性因子と，背景音を抑圧聴取する背景音因子の 2 つの重要な因子によって支えられているとし，また，音の心理的性質を主観の面から，対象音と背景音の 2 つに分けて以下のような定義を行っている．

対象音の性質

- 主な情報を持つ

- 単数または複数個の点音源で集中的
- 方向定位性で選択聴取する

背景音の性質

- 対象音のある場所の情報を持つ
- 点音源 (ノイズ) 及び空間的な音 (反響) で分散的であり、拡がりを持つ
- 一般的には聴取する場合不必要なものとして抑圧的に働く (逆にデッドな状態では必要なものになる) .

また、島田らは通信網のデジタル信号処理技術について触れ、将来目指すべき「臨場感」を伝える音通信において、一つの仮想共有空間を生成又は再生するために背景音によって音の拡がり感を制御する技術が必要であると述べている [44] .

しかし、これらの技術的要求に対して、背景音の具体的なレンダリング方法に関する研究は現在までに極めて少なく、レンダリング手法の確立、及びその実装は極めて重要な事項であると言える .

3.3 背景音レンダリング手法の検討

本研究では物理的なアプローチではなく、知覚的アプローチによる、新しいアルゴリズムを設計し、忠実性を高める手法の開発を目指す . ここで、吉田らは、背景音を「対象音のある場所の情報を持つ」と定義しているが、現時点において、対象音のある空間すべての物理現象を忠実にレンダリングするのは困難である . そこで本研究では、一般的な環境下で忠実性を向上させる効果を持つ背景音の作成を目標とし、レンダリングする空間の情報や聴取者の位置情報に依存しない背景音の作成を目指す .

3.3.1 背景音に使用する音源について

以降の検討において背景音に様々な処理を加えるため、背景音源として使用する音源は、位相特性や振幅特性が柔軟に変更可能であることが必要である . そこで、ホワイトノイズ (白色雑音) を基にすることとする . 通常ホワイトノイズは、一般的に我々が耳にする背景音に比べて相対的に高域成分のレベルが強く、背景音としては不自然に感じてしまうという問題点があるため、得られるスペクトルに何らかの補正を行い、自然な背景音に補正する必要がある .

Hoth はアメリカの電話加入者宅の騒音について調査分析し、Hoth スペクトルと呼ばれるスペクトルを提唱した [45] . このスペクトルは、CCITT (Comite Consultatif

International Telegraphique et Telephonique : 国際電信電話諮問委員会) の明りょう度試験法の勧告の中に妨害雑音の基準スペクトルとして採用されている．我が国においても，三浦らによって東京都の電話加入者について同様の調査が行われ，住宅・事務所・商店・工場の業種別平均値が求められた [46]．両者の結果は，オクターブあたり 5 dB 減衰するガウス雑音とほぼ同等とみなすことができ，高周波帯域の一部を除いて主要帯域ではかなりの類似が見られる．しかし，この両者の結果はあくまでも屋内の騒音スペクトルのデータであり，屋外空間などを含めた結果ではない．本研究では，屋内，屋外空間を問わずに適用可能な背景音を目的とするため，この結果をふまえた上で，スペクトル補正についての再検討を行う．そこで，まず背景音となりうる音の周波数特性を比較し，最適なパラメータについて検証する．ここでは音源に，様々な音を収集しデジタルな音源ファイルとして収録された音源集である SMILE ライブラリ [47, 48] より，屋内の環境音，屋外の環境音，換気扇の音を使用した．なお，ここではクラクションの音や，音声などの非定常的な音の影響が少なく，比較的定常的な音源を選んでいる．図 3.1 にそれぞれの測定環境の写真及び周波数スペクトルを示す．図はサンプリング周波数 48 kHz, 4.8×10^6 サンプルのデータをそれぞれ 1024 ポイント (21.3 ms) の時間フレームで切りだした波形の周波数特性の平均を示す．また，振幅特性は，それぞれ $f = 100\text{Hz}$ の値を 0 dB として正規化している．

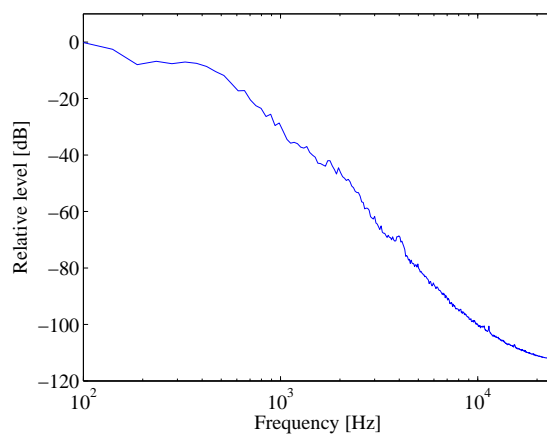
これらの周波数スペクトルを比較すると，いずれも高周波数域に向かい減衰する傾向があることが見て取れる．そこで，使用する背景音も同様の特性を持たせることを目指し，オクターブあたりのエネルギー密度をパラメータとして，ホワイトノイズをフィルタリングし，その周波数特性を比較した．式 (3.1) にしたがってフィルタを作成し，ホワイトノイズとの合成を行う，なお，式 (3.1) において， $\alpha = 0$, $\alpha = 1$, $\alpha = 2$ とすると，合成後のデータはそれぞれ，ホワイトノイズ，ピンクノイズ，レッドノイズ (ブラウニアンノイズ) になる．

$$S(f) \propto \frac{1}{f^\alpha} \quad (3.1)$$

式 (3.1) の α の値を変化させ，フィルタリング後の算出されたカラーノイズを聴き比べた結果， $\alpha = 2.0$ のときが最もこれらの背景音に近い特性を示した．そこで，以降の検討において基準となる背景音源として，このレッドノイズを使用する．また，この時の周波数スペクトルと実測の背景音 (屋外) との比較を図 3.2 に示す．高域に同程度の減衰が見られるのがわかる．



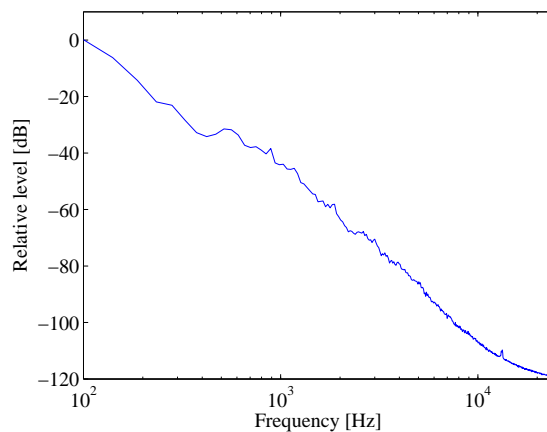
(a.1) 測定環境 1 (屋内の環境音)



(b.1) 周波数スペクトル 1



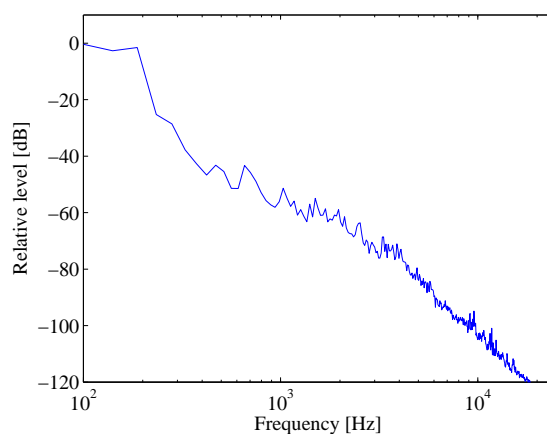
(a.2) 測定環境 2 (屋外の環境音)



(b.2) 周波数スペクトル 2



(a.3) 測定環境 3 (換気扇の音)



(b.3) 周波数スペクトル 3

図 3.1: 各背景音の測定環境と周波数特性

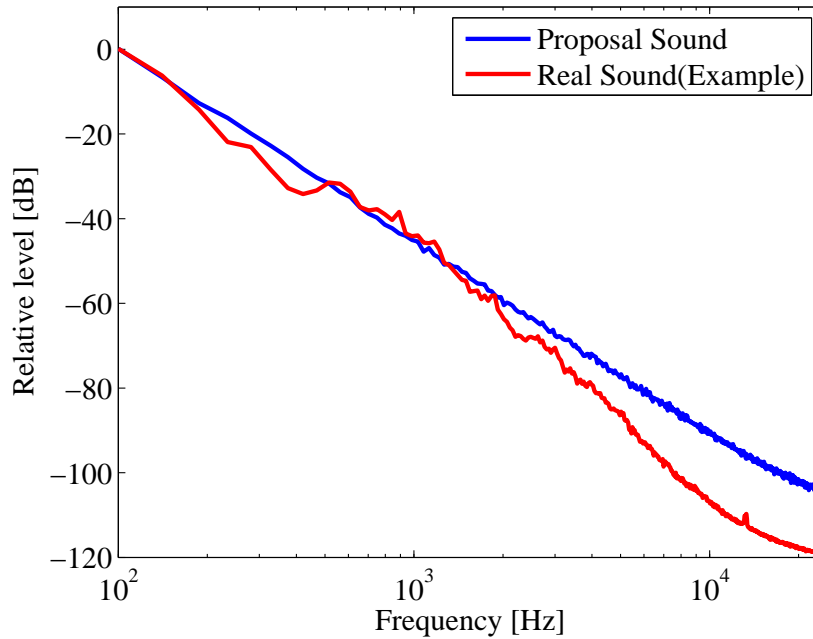


図 3.2: 提案する背景音源 ($\alpha = 2.0$) と実測の背景音 (屋外の環境音) の比較

3.3.2 レンダリング手法の検討

背景音の付加手法において、最も単純なものとして、前述の背景音源を対象音にモノラル音として加算し、そのまま出力する方法が挙げられる。この方法は BGM (Back Ground Music) と同じ付加方法であり、計算負荷が極めて少なく、背景音の作成も容易であるという利点があるが、この方法では、背景音が頭内定位してしまう。一般的に、聴覚ディスプレイにおける対象音は HRIR (Head-Related Impulse Response) が畳み込まれたものであり、頭外に定位する。そのため、耳に入ってくる音に頭内定位するものと頭外定位するものが混在してしまい、聴感上非常に不自然な音となってしまう。したがって、忠実性の高い音空間を構築しようとした場合、背景音は頭外定位させる必要があると考えられる。

この問題を解決するために、背景音源に HRIR を畳み込み、それを頭部周囲に配置する手法を取った。背景音は、水平方向に 0 度から 350 度まで 10 度刻み (図 3.3(a))、垂直方向に -30 度から 30 度まで 10 度刻み (図 3.3(b)) で計 252 点、それぞれ頭部中心から 1.5 m の位置に配置した。ここで、仰角 -40 度以下の HRIR は胴体や測定装置の影響が大きく、他の点に比べて正当性が低いと考えられる。そこで本検討で作成する背景音には、仰角 -40 度以下のデータ、及び対称性を保つために仰角 40 度以上のデータを除くこととする。以降この手法を頭部周囲配置型背景音と呼ぶ。

この背景音源群には、それぞれの角度に応じた HRIR を畳み込んでいるため、す

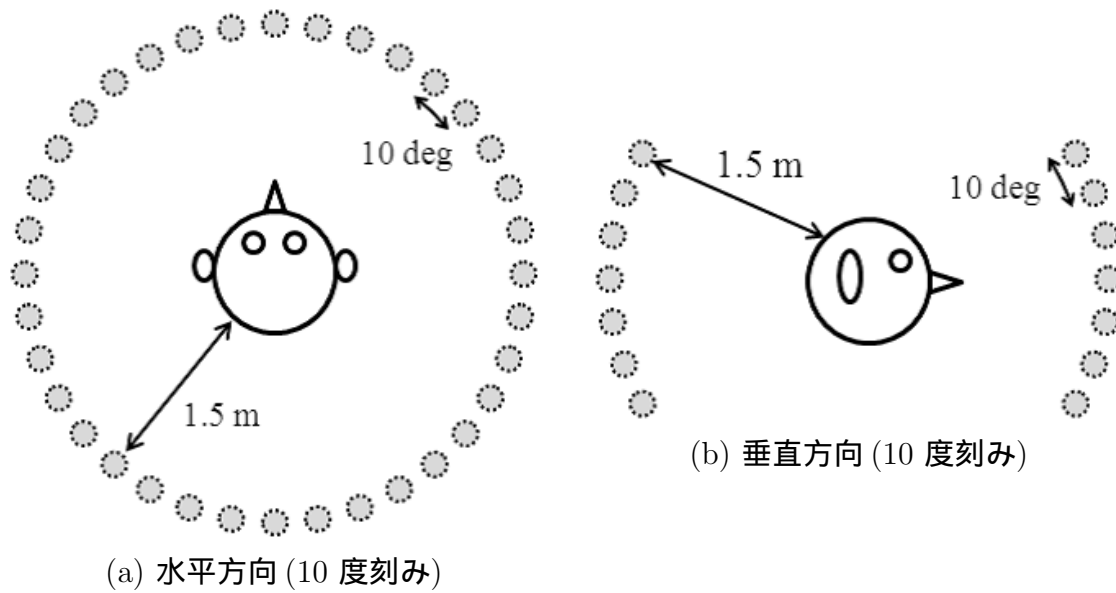


図 3.3: 頭部周囲配置型背景音における背景音源の擬似的な配置位置

すべての音が頭外定位すると考えられたが，実際に作成した背景音を聴取した結果，音が頭内で定位した．原因として，両耳に入る信号がほぼ同一であったことが考えられる．ISO 3382:1997(E)[49] によれば，式 (3.2) で与えられる正規化両耳間相互相関係数 (Inter-Aural Cross-correlation Coefficients: IACC) が 0 に近いほど，人が感じる「音の広がり感」は増し，0.4 以下であれば無指向感の関係から「臨場感」が形成され，逆に 0.75 以上であれば音は頭内定位するとされている．

$$\phi_{lr} = \frac{\Phi_{lr}(\tau)}{\sqrt{\Phi_{ll}(0)\Phi_{rr}(0)}}$$

$$\text{IACC} = \max(|\phi_{lr}(\tau)|) \quad |\tau| \leq 1\text{ms} \quad (3.2)$$

前述の同一背景音源群に対して，式 (3.2) を用いて IACC を計算した結果，0.98 程度となった．このため，作成した背景音に音の広がりを与えることができず，音が頭内に定位してしまったものと考えられる．そこで，非常に長い周期と均等性を持つ擬似乱数生成アルゴリズムである Mersenne twister[50] を用いて，あらかじめ 252 点のすべての背景音源の位相がそれぞれランダムになるように背景音源を作成し，これらにそれぞれ適切な HRTF を畳み込む．これにより，頭外定位する背景音を作成した．また，この無相関背景音源群に対して，同様に IACC を計算した結果，0.30 程度となり，音の広がり感を実現することできた．

表 3.1: 実験に使用する PC の仕様

Operating System	Microsoft Windows Vista Business
CPU	Core 2 Duo 4 1.06 GHz
Memory size	2 GB

3.4 主観評価に基づくレンダリングアルゴリズムの評価

本研究において，背景音は知覚的アプローチを含めて作成されたものであるため，物理信号によるレンダリングアルゴリズムの客観的な性能評価は困難である．そこで，聴取実験による主観評価を行い，背景音の付加，及びそのレンダリング手法の違いが，我々の知覚に対してどのような影響を与えるのかについて調べた．

そもそも，我々が実空間の様子を判断する際には，視覚情報を頼りにして認識する場合が多く，日常生活の中で音響情報のみによる判断を行うことは少ない．したがって，音響情報のみによるレンダリング効果の判断について考えると，過剰に複雑なタスクを要求せず，できるだけ単純なタスクによる判断が望ましいものと考えられる．そこで，本研究では，聴取者に「どちらがより実際聞いている音，すなわちリアルな音に近いか」という一点のみを判断させた．ここで，「実際に聞いている音」が具体的に何を示すのかは，各聴取者に自由に想定させた．これは，本研究では，どのような環境下においても忠実性を向上させる効果を持つ背景音の作成を目標とするためである．

実験方法

本実験の系を図 3.4 に示す．また，実験に使用する PC の仕様を表 3.1 に示す．

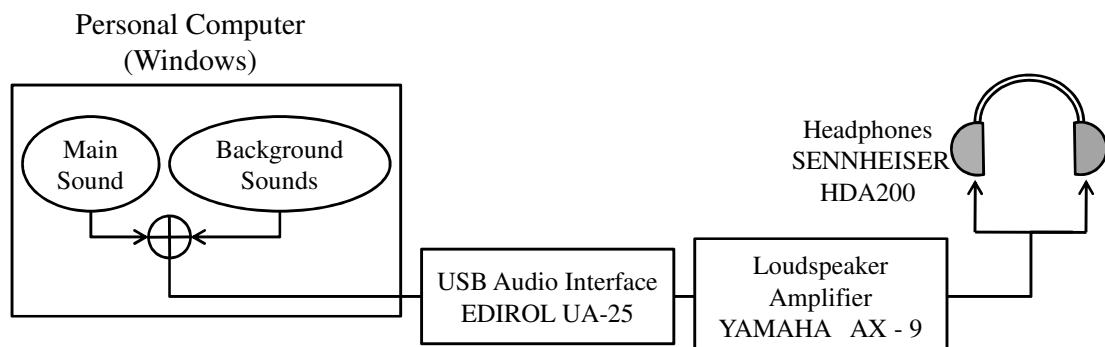


図 3.4: 実験系

実験は防音室内で行った．実験に用いる刺激の対象音には，SMILE ライブラリ [47, 48] から選出した 4 種類 (チェロ，フルート，オーボエ，バイオリン) の無響室録音された演奏音に対して，ダミーヘッド (高研 SAMRAI) の HRIR を畳み込んだものを使用した．実験刺激の長さはそれぞれ 10 秒程度であり，時間当たりのエネルギーが等しくなるように正規化を施した．本実験では背景音が知覚に与える影響を検証するために，対象音に対するレンダリング処理は HRIR による音源方向特性の付加のみとし，SiFASo に実装されている反射音やドップラー効果などのレンダリング処理は施さないこととした．

また，実環境の背景音の影響は極力排除するために，遮音性に優れる密閉型のヘッドホンを使用することとした．本実験では周波数特性やダイナミックレンジに優れる SENNHEISER の HDA200[51] を使用した．

対象音と背景音の耳元におけるパワー比，すなわち SNR(Signal to Noise Ratio) は， $SNR = -5 \text{ dB}, 0 \text{ dB}, 5 \text{ dB}, 10 \text{ dB}$ の 4 種類とした．したがって，一つの背景音対について，16 回の判断が成されることになるため，本実験の総試行数は，計 ${}_3C_2 \times 16 = 48$ 試行となる．なお，レンダリング手法による影響の差を検討するために，対象音源の種類，及び SNR については条件としないこととした．したがって，実験で使用する刺激対は，レンダリング手法が異なり，対象音源の種類，及び SNR が同じであるものについて一対比較を行った．

実験に参加した被験者は正常な聴力を持つ成人男女 8 名である．被験者には同じ対象音，SNR で，レンダリング手法の異なる 2 種類の音を聞かせ，「どちらが普段実際に聞いている音に近いか」を強制 2 肢選択で判断させた．「実際に聞いている音」は，各聴取者に自由に想定させたこととした．これは本研究で目指す背景音に，どのような環境下においても忠実性を向上させる効果を持たせるためである．

音圧校正は音源に対象音のみの刺激を基準として， $L_{Aeq} = 60 \text{ dB}$ となるように設定した．

実験 1-1：背景音の付加効果の検証

本実験においては，まず背景音の付加そのものの効果を検証するために，実験刺激として，対象音のみの刺激 (以下，None 条件，図 3.5) と，背景音を単純加算した刺激 (以下，Simple 条件，図 3.6) の比較を行った．また，評価のリファレンスとして，対象音のみの刺激に，ダミーヘッドによってバイノーラル録音された実空間の背景音を付加した刺激 (以下，Real 条件，図 3.7) を使用した．

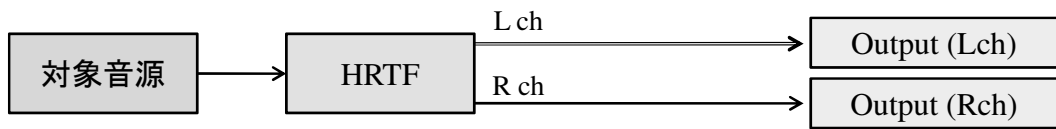


図 3.5: None 条件の概念図

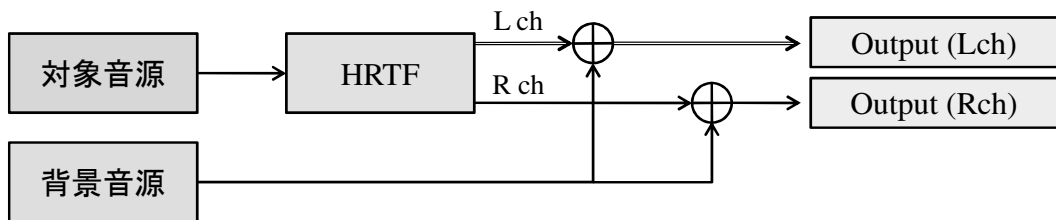


図 3.6: Simple 条件の概念図

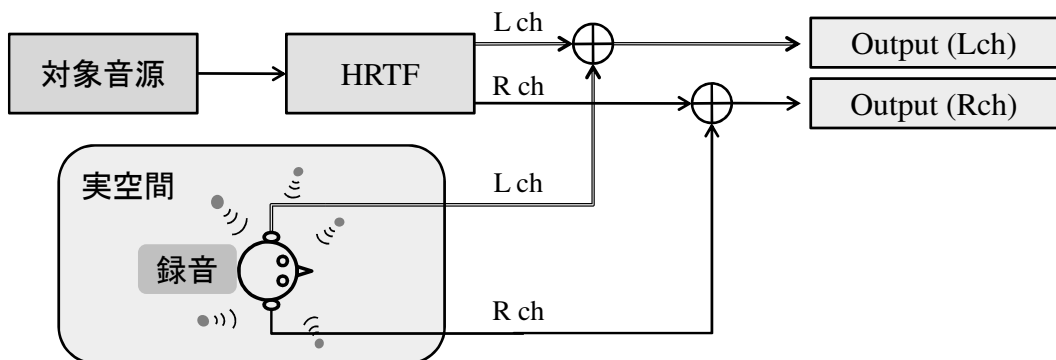


図 3.7: Real 条件の概念図

結果

実験結果を図 3.8 に示す．図 3.8 は実験 1-1 における，参加者全員の各レンダリング手法に対する平均の心理距離尺度値を表わしている．分散分析の結果，条件の主効果が有意であった ($F(2, 20) = 31.261, p < .001$)．多重比較 (Bonferroni の不等式に基づく多重比較法 [52]) を行った結果，Simple 条件，及び Real 条件は，None 条件よりも有意に値が高かった．また，Real 条件は，Simple 条件よりも有意に値が高かった．

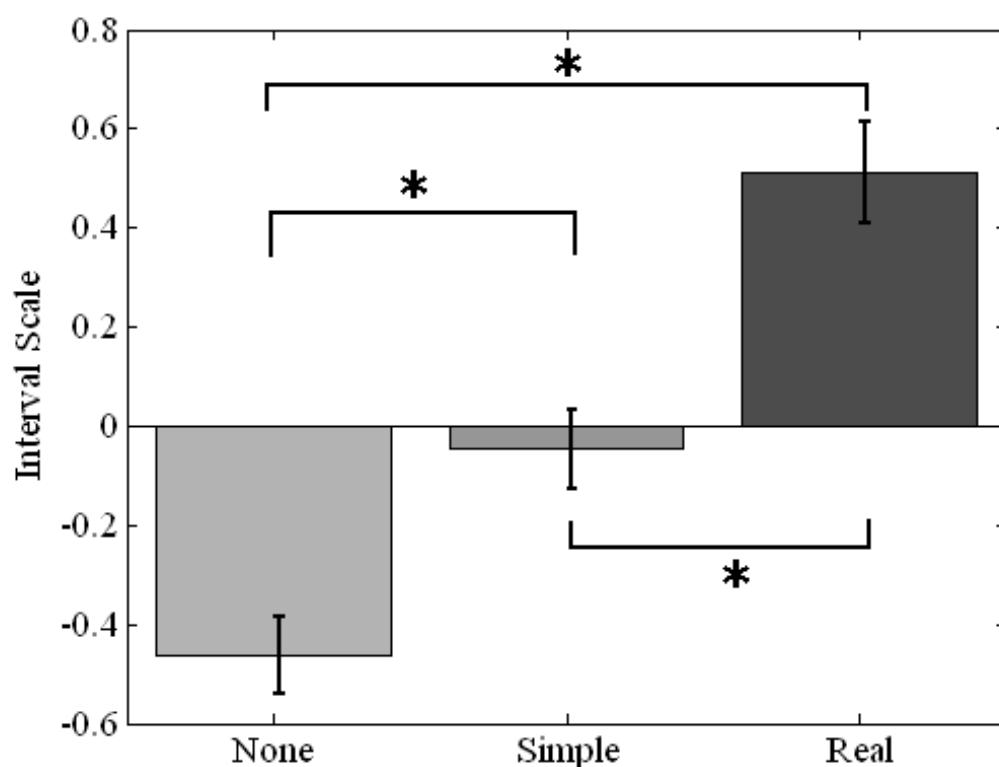


図 3.8: 実験 1-1 : 各レンダリング手法に対する心理距離尺度値 (参加者全体の平均，* : $p < .05$)

また，結果を SNR ごとに整理し，全被験者の判断比率の平均を取ったものを全組み合わせについて，それぞれ図 3.9，図 3.10，図 3.11 に示す．

実験 1-2 : 周囲配置型背景音の有用性の検証

次に，頭外に定位する周囲配置型背景音の付加効果を検証するために，実験刺激として，対象音に背景音を単純に加算した刺激，周囲配置型背景音を加算した刺激 (以

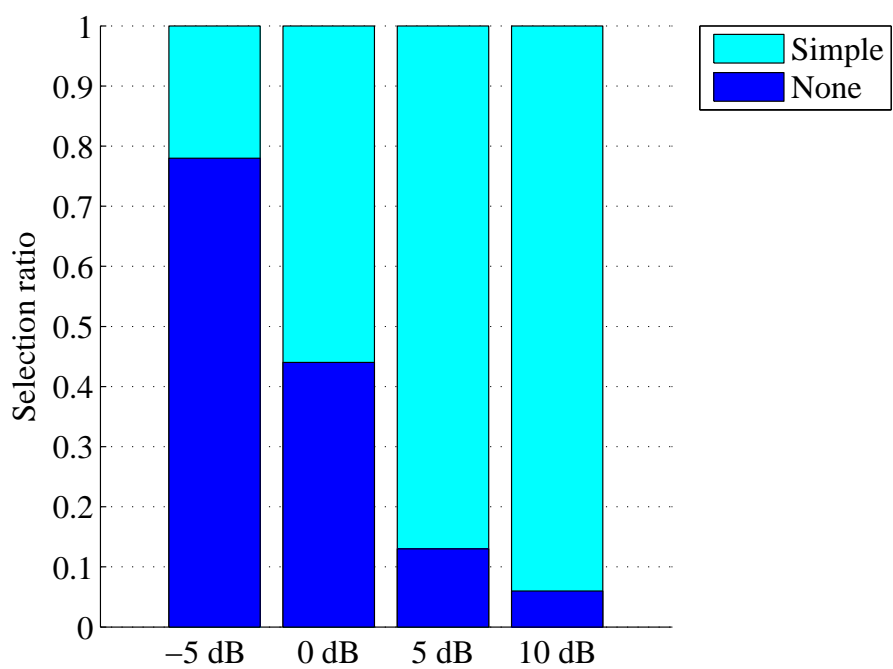


図 3.9: 実験 1-1 : SNR ごとの平均判断比率 (None 条件 vs. Simple 条件)

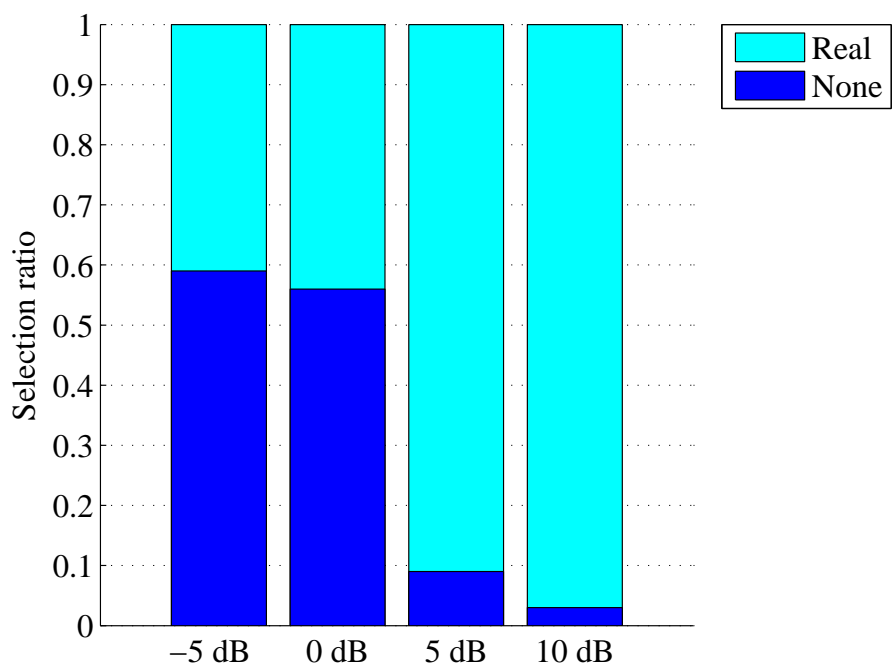


図 3.10: 実験 1-1 : SNR ごとの平均判断比率 (None 条件 vs. Real 条件)

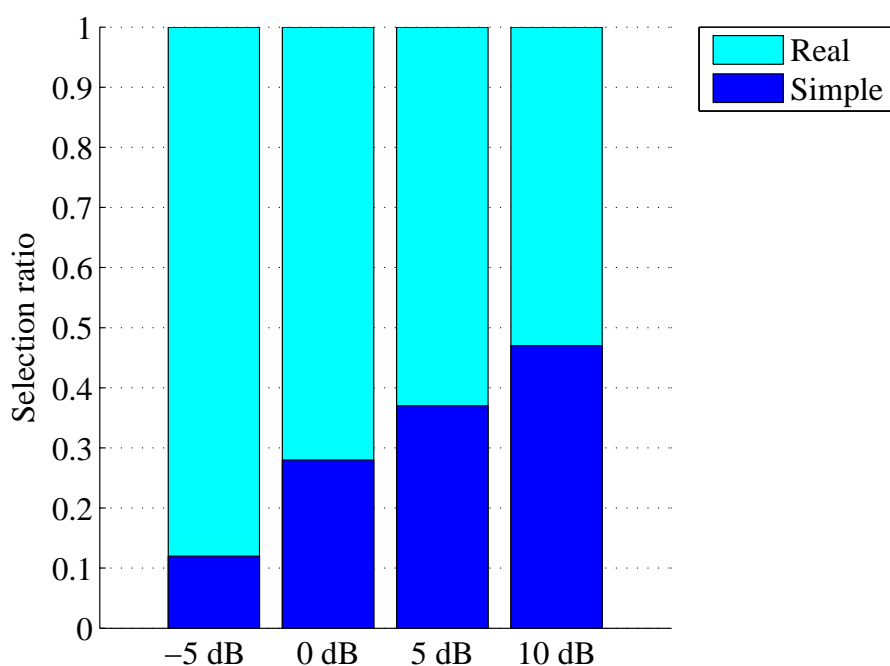


図 3.11: 実験 1-1 : SNR ごとの平均判断比率 (Simple 条件 vs. Real 条件)

下, Surround 条件 (図 3.12), バイノーラル録音された背景音を加算した刺激を使用し, これらの刺激を用いて一対比較による評価を行った。

実験 1-2 : 結果

実験結果を図 3.13 に示す。図 3.13 は実験 1-2 における, 参加者全員の各レンダリング手法に対する平均の心理距離尺度値を表わしている。分散分析の結果, 条件の主効果が有意であった ($F(2, 23) = 12.067, p < .001$)。多重比較 (Bonferroni の不等式に基づく多重比較法) を行った結果, Real 条件, 及び Surround 条件は, Simple 条件よりも有意に値が高かった。また, Real 条件と Surround 条件の間に有意な差は見られなかった。

また, 結果を SNR ごとに整理し, 全被験者の判断比率の平均を取ったものを全組み合わせについて, それぞれ図 3.14, 図 3.15, 図 3.16 に示す。

結果に関する考察

実験 1-1 において, Simple 条件が, None 条件よりも有意に高い値を示したことから, 背景音を加えることで忠実性が向上することが示された。しかし, Simple 条件

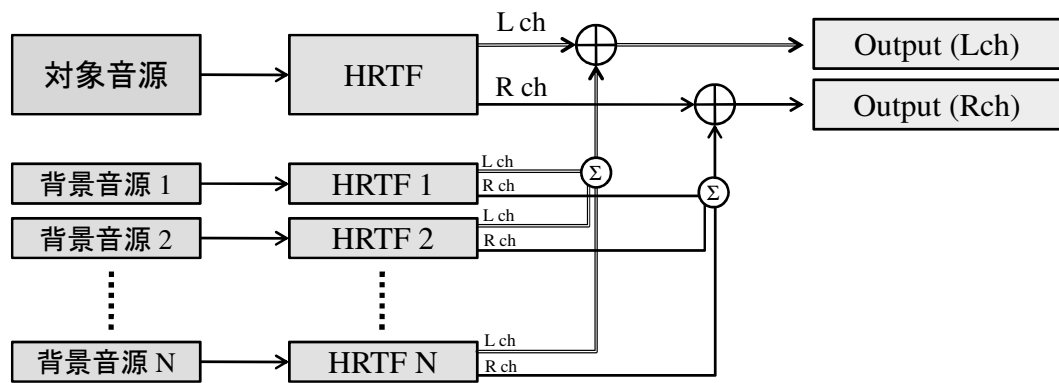


図 3.12: Surround 条件の概念図

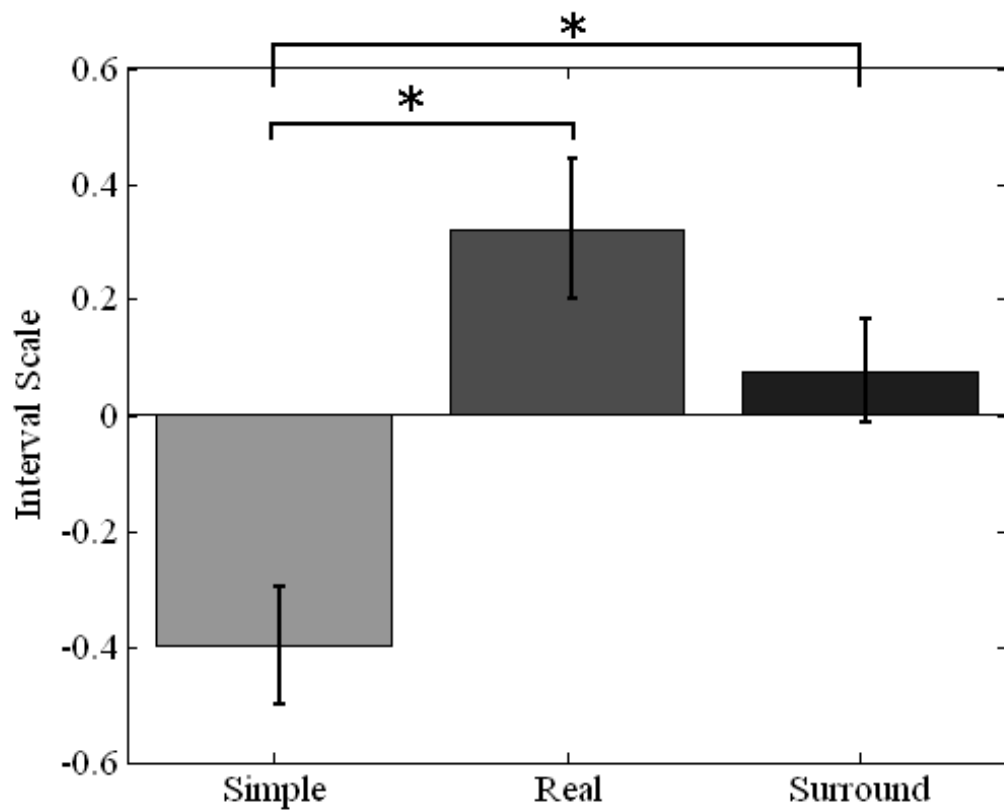


図 3.13: 実験 1-2 : 各レンダリング手法に対する心理距離尺度値 (参加者全体の平均, * : $p < .05$)

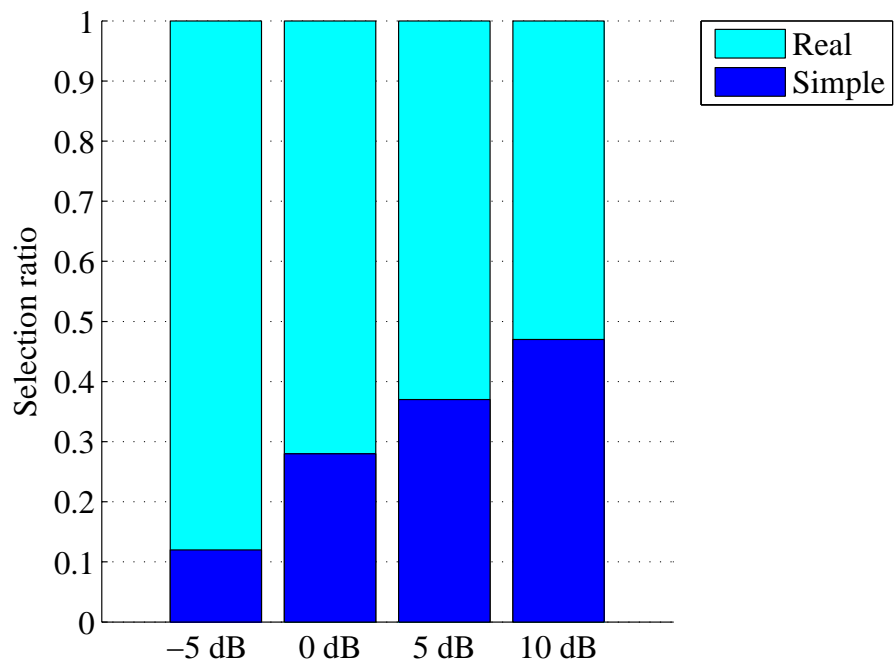


図 3.14: 実験 1-2 : SNR ごとの平均判断比率 (Simple 条件 vs. Real 条件)

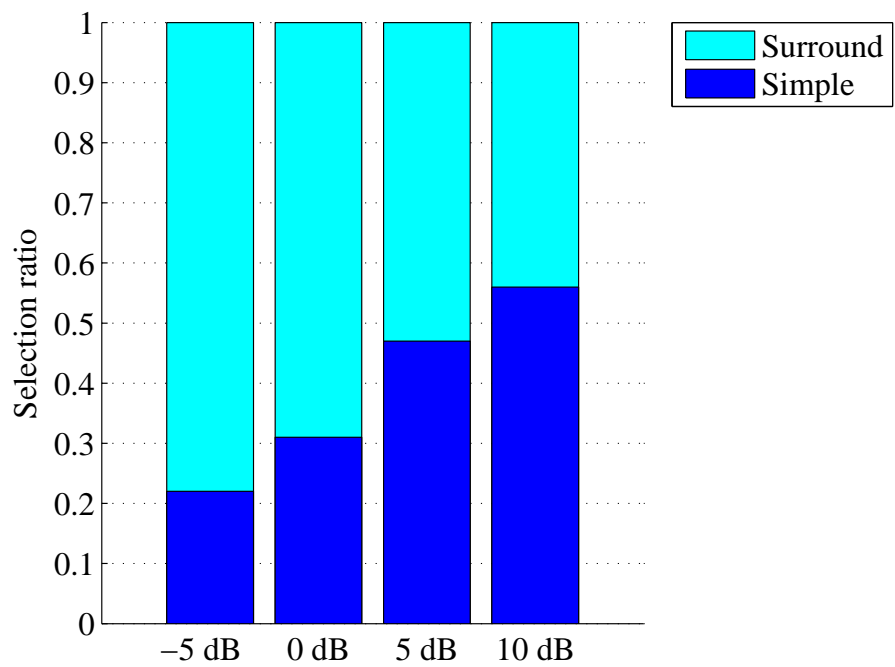


図 3.15: 実験 1-2 : SNR ごとの平均判断比率 (Simple 条件 vs. Surround 条件)

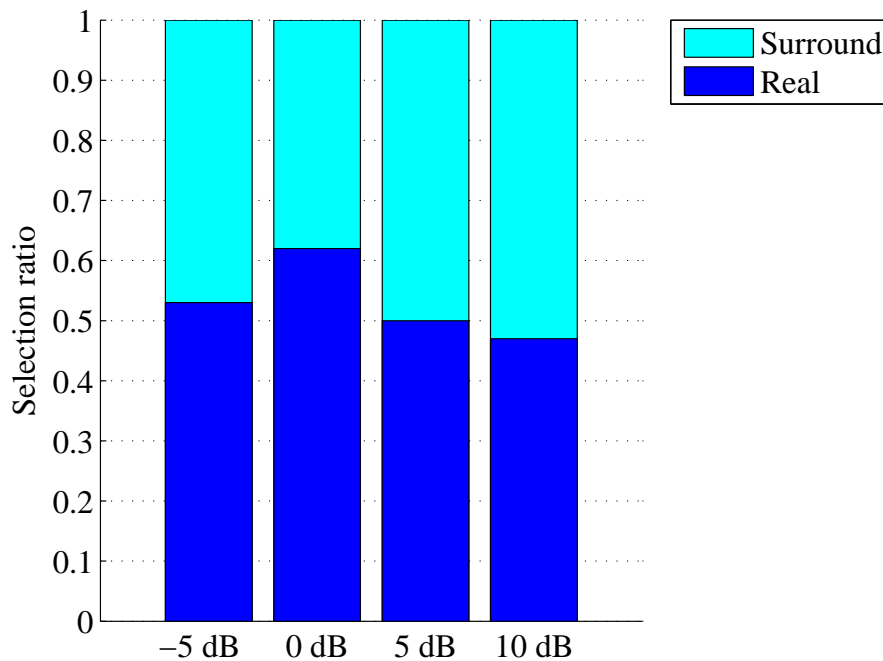


図 3.16: 実験 1-2 : SNR ごとの平均判断比率 (Real 条件 vs. Surround 条件)

は Real 条件よりも有意に値が低く，単純に背景音を加算するだけでは，実測の背景音の忠実性に及ばないことがわかる．これは，単純加算した背景音は頭内に定位するため，不自然に感じた被験者が多かったためと考えられる．

また，SNR ごとの結果を見ると，SNR が低い(背景音のレベルが相対的に高い)場合，None 条件を選択する比率が高い傾向が見られた．実空間において想定する音圧レベル以上の背景音を付加した場合，逆に忠実性が低下することが示唆された．特に，図 3.9 の SNR = -5 dB の場合においてはこの傾向が顕著に見られた．単純加算した背景音である Simple 背景音は，頭内定位するため，不自然さが際立ったものと考えられる．この傾向は，図 3.11 においても同様に見られ，頭外定位する Real 背景音と，頭内定位する Simple 背景音は SNR が低いほど差が顕著に現れた．

実験 1-2 において，Surround 条件が Simple 条件よりも有意に高い値を示したことから，背景音を頭外定位させることによって忠実性が向上することが示された．加えて，Real 条件と Surround 条件の間に有意差が見られなかったことから，周囲配置型背景音は実測の背景音を使用した場合と同程度の忠実性を与えることが可能であると考えられる．また，SNR ごとの結果においても Real 条件と Surround 条件の間に特定の傾向は見られなかった．

実験後の被験者にアンケートを取った結果，対象音と背景音 (アンケート中では雑

音と表記されていた) が分離していて不自然に聞こえるという意見が多く見られた。原因として、視覚情報が与えられなかったために、実空間のイメージが困難であったことや、対象音に無響室において測定されたドライソースに HRTF の畳み込み処理のみを施した音を使用したために、背景音があるにも関わらず、不自然にはっきりとした音になっていたこと、さらに、多くの被験者が屋外環境をイメージしていたために、通常屋外で聞く機会の少ない楽器音に違和感を感じたことなどが理由として考えられる。これらは主に対象音に関する問題点であるが、本研究では対象音に特別な処理を加えずに、背景音のみで忠実性を高めることを目指している。そのため、処理を加えるのは可能な限り背景音のみとしたい。

実空間と今回の実験環境の大きな違いとして、頭部運動に出力音が連動していないことが挙げられる。対象音は音源方向が認識しやすいため、頭部運動に敏感であるが、背景音は音源方向が認識しにくいいため、頭部運動に伴う定位方向の変化が対象音よりも鈍くなると予想される。そこで、「頭部運動に連動する定位方向の差が背景音の感じ方に影響する」という仮説を立てた。次章では、上記の聴取者の頭部の運動に対応した聴覚ディスプレイを使用し、背景音の効果についての詳細な検証を行う。

3.5 まとめ

本章では背景音の付加手法に関する検討を行った。既存の聴覚ディスプレイシステムは対象とする音のみを出力する仕様となっているため、出力される音は、非常に静かな録音スタジオでコンサートを聴いているかのような非常に不自然な音となっていた。そのため、本研究では出力に背景音を付加することで忠実性の向上を図った。

まず、背景音として使用する音の検討を行った。本研究では屋外空間を含めた一般的な状況下に適用可能な背景音の作成を目的としているが、背景音スペクトルとして一般的に知られる Hoth のスペクトルや三浦らのスペクトルは室内において測定されたものである。そこで、屋外の背景音を含めた様々な背景音のサンプルをもとに理想的な背景音源についての検討を行った結果、ホワイトノイズを $S(f) \propto 1/f^{2.0}$ で与えられる周波数特性でフィルタリングすることで得られる信号を背景音源として使用する。この背景音源を出力信号に単純加算しただけでは、背景音が頭内定位してしまうという問題があったため、252 個の無相関な背景音源群に、それぞれ異なる角度の HRIR を畳み込むことで、頭外定位する背景音を作成した。

次に背景音の付加効果を検証するために主観評価による実験を行った。その結果、提案する背景音源を出力音に単純加算するだけでも忠実性が向上することが示された。また、単純に加算しただけでは実空間においてバイノーラル録音した背景音には及ばないものの、頭外定位する背景音を使用すれば、実測の背景音と同程度の忠実性

を与えることが可能となることが示された．

しかし，実験後の内省報告において，実験刺激に不自然さを感じたという意見が多数見られた．次章で頭部運動の寄与を検証する

第4章 頭部運動を考慮した背景音のレンダリング

4.1 はじめに

前章の検討により，聴取者の頭部運動に出力が感応しない静的な状況下においては，仮に背景音を加えたとしても，ある一定以上の忠実性を与えることはできないのではないかと予想された．そこで，本章では頭部運動と背景音の相乗効果による忠実性への寄与を検証する．

本章では刺激音の提示に，矢入らの開発した頭部運動感応型聴覚ディスプレイ [13] を使用する．矢入らの聴覚ディスプレイは，HRIR による方向特性の付加に特化したシステムであるため，システム遅延が非常に小さく，HRIR の空間，時間補完精度も高い．そのため，背景音の影響について検証するには最適なシステムであると考えられる．

まず，前章で提案した Surround 背景音を頭部運動に感応させることで忠実性の向上を目指す．Surround 背景音を構成する背景音源一つ一つに HRIR をリアルタイムで畳み込むことは不可能であると考えられるため，本研究では頭部角度に対応した Surround 背景音をオフラインで作成しておき，リアルタイムで出力される位置データに応じた適切なデータを出力することで Surround 背景音の頭部運動への感応を実現する．

しかし，Surround 背景音を構成する背景音源間の違いは位相のみであり，頭部運動によって出力が変化したとしても，その違いを認識することは非常に困難である．この問題を解消するために，各背景音源間に振幅の差を与える．これにより，頭部運動による変化の違いを聴取者が認識しやすくなると考えられる．

これらの検討事項を聴取実験によって検証し，結果をもとに忠実性を高めるための背景音の条件について考察する．

4.1.1 耳入力音信号の動的変化の効果

日常生活において，頭部を完全に固定していることはない．外を歩く場合など，耳に入ってくる音の相対位置は常に移動していることになる．また，仮に聴取者本人が

完全に静止していたとしても，人の声，自動車の音などの，音源が移動する場合も多くある．すなわち，普段音を聞く場合には，聴取者や音源が動くことにより両耳における音信号が変化することがほとんどであり，音信号を動的に変化させることが聴取者の感じる忠実性を向上させることは明らかである．

耳入力信号の動的変化が聴覚知覚に与える影響については，現在まで主に音像定位について検証が行われてきた．例えば Wallach [53] は，音信号の両耳間差，両耳間レベル差の動的変化に着目し，頭部運動による音像の仰角，前後判断の定位メカニズムについて考察した．音源が水平面から離れるにつれて，頭部を左右に回転させた際の両耳間差の変化率が小さくなるが，彼はこのことが定位の手がかりであると仮定し，水平面上にあるスピーカで仰角方向の音源を模擬した．その結果，頭部の回転による音の動的変化は，前後判断及び仰角判断において手がかりとなることを示した．Wallach はまた，音源の移動や聴取者の動きが音像定位の曖昧さを解決する上で有効であると述べている [54]．また近年では，実音源聴取時のみならず，聴覚ディスプレイなどによる仮想音源の聴取時の定位精度にも影響を及ぼすことが検証されている (例えば [55, 56, 57]) ．

しかし，これらの頭部運動に伴う耳入力信号の動的変化において，具体的に忠実性を向上させる要因については不明な点が多い．この要因を解明し，出力される音に付加することができれば，忠実度の高いシステムの実現が期待できる．

4.1.2 頭部運動と背景音についての仮説

前章の検討により，聴取者の頭部運動に出力が感応しない，静的な状況下においては，背景音の付加による忠実性の向上が少ないと考えられる．

対象音に初期反射や残響などの音響効果を付加することで，提示する音空間の忠実性を高めることは可能であると考えられるが，本検討では背景音のみの処理によって，音空間全体の忠実性を高めることを目指す．背景音の条件のみを考えたとき，実空間で音を聴取する環境と第 3 章の実験環境の大きな違いとして，頭部運動に出力音が感応していないことが挙げられる．

吉田らの定義 [43] を用いれば，対象音は「単数または複数個の点音源で集中的」であり，かつ「方向定位性」を持つ．すなわち対象音は音源方向を認識しやすいものであり，頭部運動に伴い定位方向が敏感に変化すると考えられる，一方背景音は，「空間的な音で分散的」であり「拡がり」を持つものであると定義されている．したがって，背景音は音源方向が認識しにくい，認識不可能であり，頭部運動に伴う定位方向の変化が対象音よりも鈍くなると予想される．そこで，吉田らの定義をもとに「頭部運動に感応する対象音と背景音の定位方向変化の差が音空間の感じ方に影響する」

という仮説を立てた．以下ではこの仮説を検証し，忠実性の向上を目指す．

4.2 実験 2-1：頭部運動と背景音の関連についての検証

頭部運動が背景音を含めた音空間の知覚に与える影響を検証するために，頭部運動に感応する背景音と，感応しない背景音の比較を行う．本章の検討を行うためには，聴取者の頭部運動に感応した音を提示する必要があるため，聴取者が自由に頭部運動を行うことが可能な頭部運動感応型聴覚ディスプレイを使用する．SiFASo も頭部運動感応型聴覚ディスプレイの一つであるが，本検討では矢入らの開発した聴覚ディスプレイ [13] を使用する．矢入らの聴覚ディスプレイは，自由空間における方向特性のみを付加した対象音を出力する聴覚ディスプレイであるため，他の聴覚ディスプレイに比べてシステム遅延が非常に小さく，更新レートも高い．また HRIR の時間補完精度，空間補完精度に優れ，また使用する位置センサの空間精度も高い．本検討では方向特性のみを付加した音を対象音として使用するため，背景音の影響を検証するためのシステムとして，現状における最適のシステムであると考えられる．

実験刺激の作成

前章の検討によって，Surround 背景音は，実空間で測定された背景音とほぼ同程度の忠実性を与えることが示された．そこで，本章においても，この Surround 背景音をベースとして背景音を作成する．

矢入らのシステムは，位置センサからリアルタイムで送られた情報をもとに，最適な HRIR を選択，あるいは補間によって算出し，対象音に畳み込むことで出力音を頭部運動に感応させている．背景音を頭部運動に感応させるためには，対象音と同様の処理を背景音に対して行えばよいが，Surround 背景音は，計 252 点の音源信号を加算したものであるため，その一つ一つにリアルタイムで HRIR を畳み込むことはほぼ不可能である．そこで，あらかじめ頭部の向きに対応した角度ごとに Surround 背景音 $S_{\theta_1}(t)$, $S_{\theta_2}(t)$, \dots , $S_{\theta_N}(t)$ を作成しておき，リアルタイムで聴取者の現在の頭部の向き θ_k を検出した後，適切な Surround 背景音 $S_{\theta_k}(t)$ を出力する．PC のメモリを考慮して，用意する背景音は，水平角 5 度刻みで 72 個，各 1 s (48,000 sample) とし，この背景音を繰り返し出力することとする．ここで，背景音を仰角方向への頭部運動にも対応させるには，現状の数倍ものメモリが必要となり，また HRIR の仰角変化による聴感上の影響は水平角方向の変化に比べて小さいため，本実験において，背景音は水平角方向のみの頭部運動に対応させることとする．

表 4.1: 実験に使用する PC の仕様

Operating System	Linux (Kernel 2.6)
CPU	Pentium 4 3.06 GHz
Memory size	2 GB
Sound Driver	OSS
Sound Card	M-Audio AudioPHILE 2496

2 パターンとし，頭部運動によって背景音知覚にどのような影響があるかについて検討する．

実験は第 3 章における実験と同様に防音室内で行い，実験に用いる刺激の対象音は SMILE ライブラリ [47, 48] から対象音には 4 種類 (チェロ，フルート，オーボエ，バイオリン) の無響室録音された演奏音に対して，ダミーヘッド (高研 SAMRAI) の HRIR を畳み込んだものを使用した．実験刺激の長さはそれぞれ 10 秒程度であり，時間当たりの音のエネルギーがそれぞれ等しくなるように正規化を行った．本実験においても対象音に対するレンダリング処理は HRIR (Head-Related Transfer Function) による音源方向特性の付加のみとした．

本実験において，SNR は $SNR = -5$ dB, 0 dB, 5 dB, 10 dB の 4 種類とした．したがって，一つの背景音対について，16 種類の音源対を比較することになる．また，カウンターバランスを考慮して，提示順序を逆の場合を加えるため，計 32 回の判断が成される．なお，レンダリング手法による影響の差を検討するために，対象音源の種類，及び SNR については条件としないこととした．したがって，実験で使用する刺激対は，レンダリング手法が異なり，対象音源の種類，及び SNR が同じであるものについて一対比較を行った．

実験に参加した被験者は正常な聴力を持つ成人男女 9 名である．被験者には同じ対象音，SNR で，レンダリング手法の異なる 2 種類の音を聞かせ「どちらが普段実際に聞いている音に近いか」を強制 2 肢選択で判断させた．「実際に聞いている音」は，各聴取者に自由に想定させたることとした．これは本研究で目指す背景音が，どのような環境下においても忠実性を向上させる効果を持つためである．

また，本来聴取時の頭部運動は，聴取者の自由に動かせることが望ましいが，聴取者がある程度頭部を大きく動かさなければ，背景音の差を認識することが困難になると予想された．そこで，本実験では，音がしている間に少なくとも一回ずつ左右を見る (頭部を動かす) ことを強制させた．Thurlow ら [59] によって，頭部を自由に動作することを許可した場合と，動作範囲を強制指示した場合に，定位精度等の差は見られないことが報告されているため，頭部運動の強制指示が聴取者に与える影響は小さ

いと考えられる．

音圧校正は $SNR = 0$ dB の Fix 条件を基準として， $L_{Aeq} = 60$ dB となるように設定した．

実験のタイムパターンを図 4.2 に示す．1 つ目の刺激音 (約 10 秒) を提示し，1 秒間の無音区間の後に，対となる刺激 (約 10 秒) を提示した．被験者は次の刺激音対の 1 つ目が提示される 5 秒後までに回答を行った．

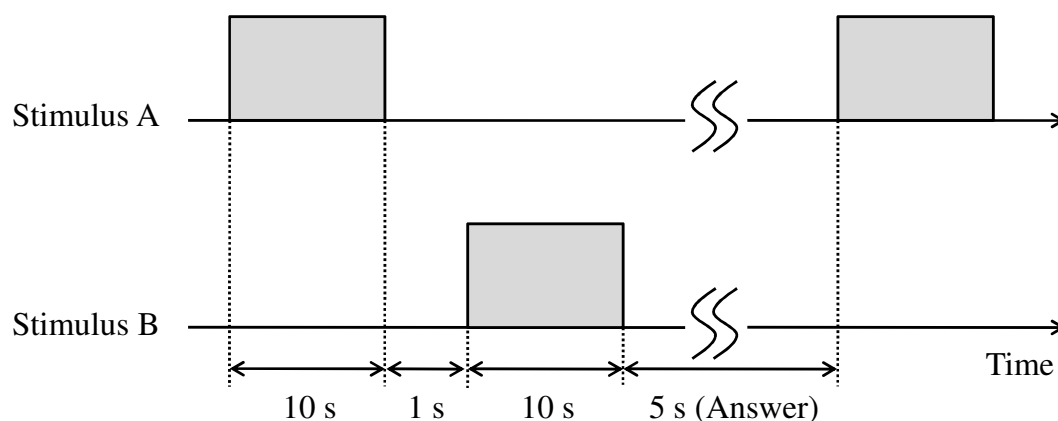


図 4.2: 実験のタイムパターン

結果

実験の結果をもとに t 検定を行ったところ，Move 条件と Fix 条件の間に有意な差はなかった ($t(8) = 0.87, p > .05$)．全被験者の判断比率を表 4.2 に示す．また，SNR ごとの全被験者の判断比率の平均を図 4.3 に示す．

考察

実験の結果，背景音を動かしたことによる差は見られなかった．また，SNR ごとに見た結果にも特定の傾向は見られなかった．原因として，Surround 背景音を構成する各背景音源の違いが，位相のみであったことが挙げられる．位相のみが異なるホワイトノイズを弁別することは非常に困難であるため，頭部の運動に Surround 背景音を感応させてもその違いを認識できなかったものと考えられる．

そこで，Surround 背景音を構成する各背景音源間に，振幅の差を持たせることとする．この処理により，Surround 背景音が頭部運動に感応していることを聴取者に認識させることが可能になり，忠実性を向上させることが可能になると予想した．

表 4.2: 実験 2-1 における全被験者の判断比率

	Rendering Method	
	Fix	Move
sub1	0.59	0.41
sub2	0.47	0.53
sub3	0.38	0.63
sub4	0.50	0.50
sub5	0.72	0.28
sub6	0.63	0.38
sub7	0.56	0.44
sub8	0.38	0.63
sub9	0.59	0.41
Average	0.53	0.47

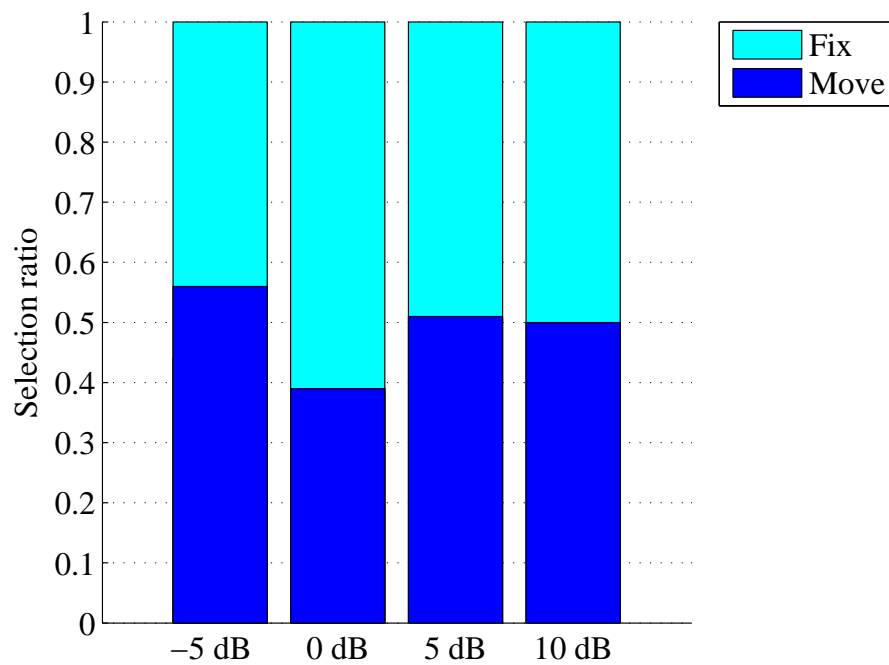
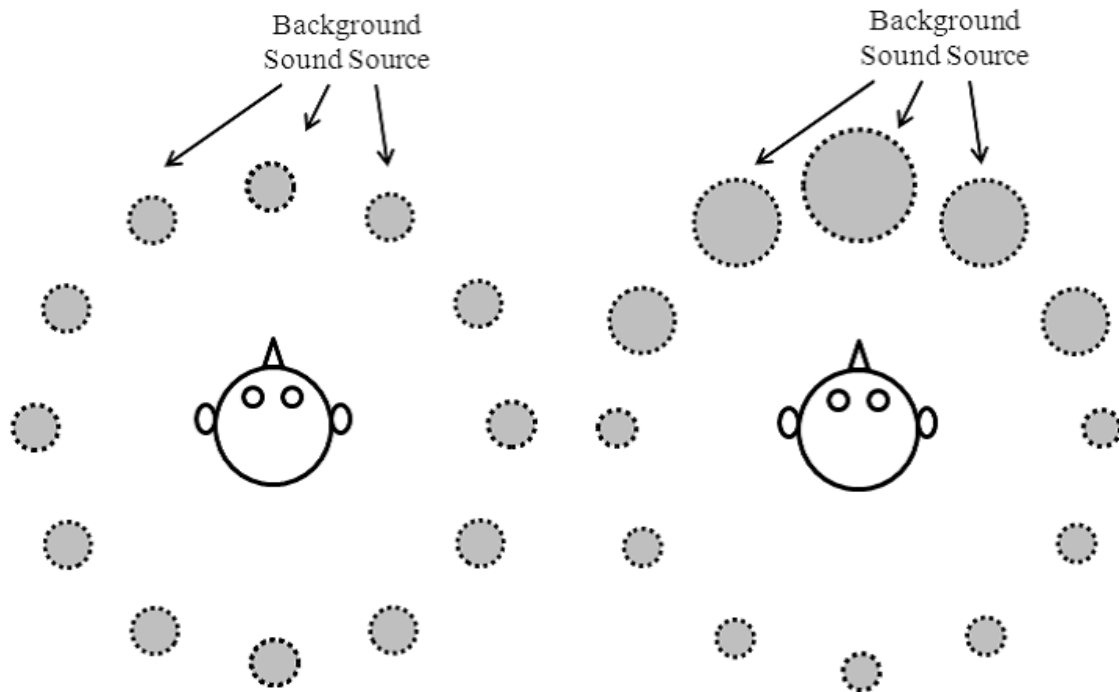


図 4.3: 実験 2-1 : SNR ごとの平均判断比率

4.3 実験 2-2：背景音群の振幅に変化を与えた場合における検証

背景音源の音圧レベルを変化させる際に，ある背景音源一点だけの音圧レベルを高めた場合，その方向に非常に鋭い定位が可能な背景音源が存在することになり，これは本章のはじめに立てた仮説に反する．そこで，本実験においては擬似的に定義した背景音源位置（角度）に対し，この位置からの角度の差に応じて振幅を変化させることで，なだらかに音圧レベルを変化させた．実験 2-1 で使用した Surround 背景音と，本実験で使用する，特定方向近辺に配置された背景音源の音圧レベルを意図的に高めた Surround 背景音の違いを，図 4.4 に示す．



(a) 実験 2-1 で使用した Surround 背景音 (b) 本実験で用いる Surround 背景音

図 4.4: 本検討で用いる Surround 背景音と通常の Surround 背景音との比較

実験刺激の作成

水平角を θ ，仰角を ϕ に対して，擬似音源方向の角度を $(\theta, \phi) = (0, 0)$ としたときに， $(\theta, \phi) = (0, 0)$ で最大値 1 を取り， $\theta = \pm 90 \text{ deg}$ または， $\phi = \pm 90 \text{ deg}$ で最小値 0，その他がすべて 0 となるような 2 次元 Hanning 窓を用いて，背景音源の振幅に変化を与える．この窓を図 4.5 に示す．

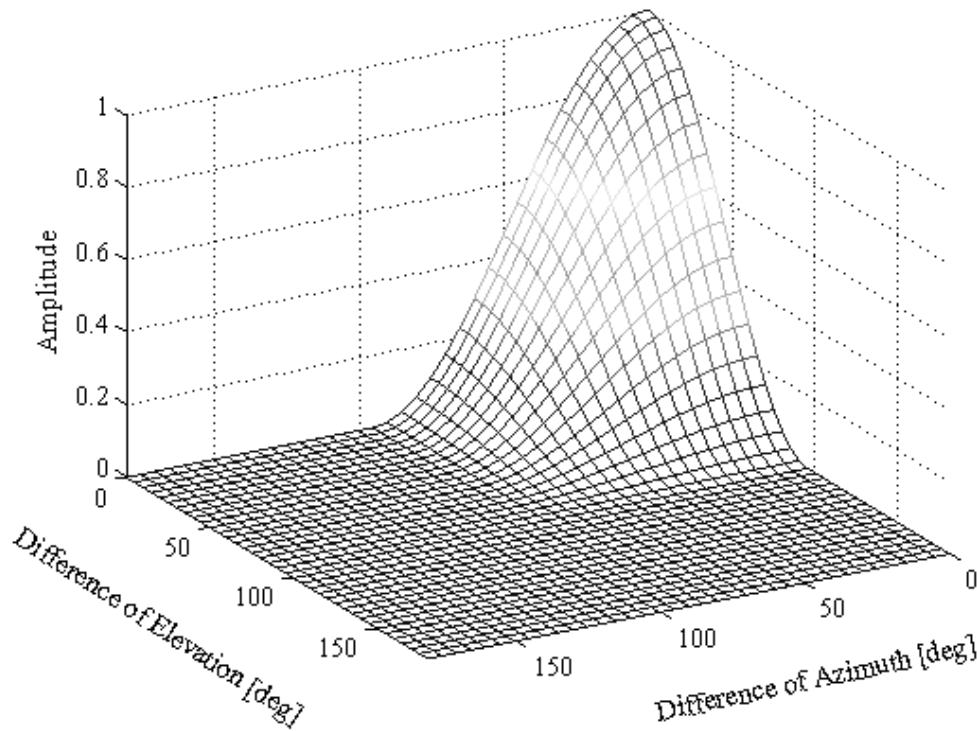


図 4.5: 用いた 2 次元 Hanning 窓

本研究では，この 2 次元 Hanning 窓の振幅を 9 倍し，すべての点に対して 1 を加えたものを使用する．したがって，最大音圧となる背景音源は最小音圧となる背景音源に比べて振幅が 10 倍となり，20 dB 高いものとなる．

また，本実験において，実験装置の構成，及び実験方法は，実験 2-1 と同様の条件とする．

結果

実験の結果をもとに t 検定を行ったところ，Move 条件は Fix 条件よりも有意に高い値を示した ($t(8) = 2.48, p < .05$) ．

全被験者の判断比率を表 4.3 に示す．また，SNR ごとの全被験者の判断比率の平均を図 4.6 に示す．

考察

この結果から，背景音を頭部運動に感応させることで忠実性が向上することが示された．また，SNR ごとの結果を見ると，SNR が低い場合，すなわち背景音の音圧レ

表 4.3: 実験 2-2 における全被験者の判断比率

	Rendering Method	
	Fix	Move
sub1	0.47	0.53
sub2	0.50	0.50
sub3	0.34	0.56
sub4	0.44	0.56
sub5	0.47	0.53
sub6	0.50	0.50
sub7	0.34	0.66
sub8	0.47	0.53
sub9	0.50	0.50
Average	0.45	0.55

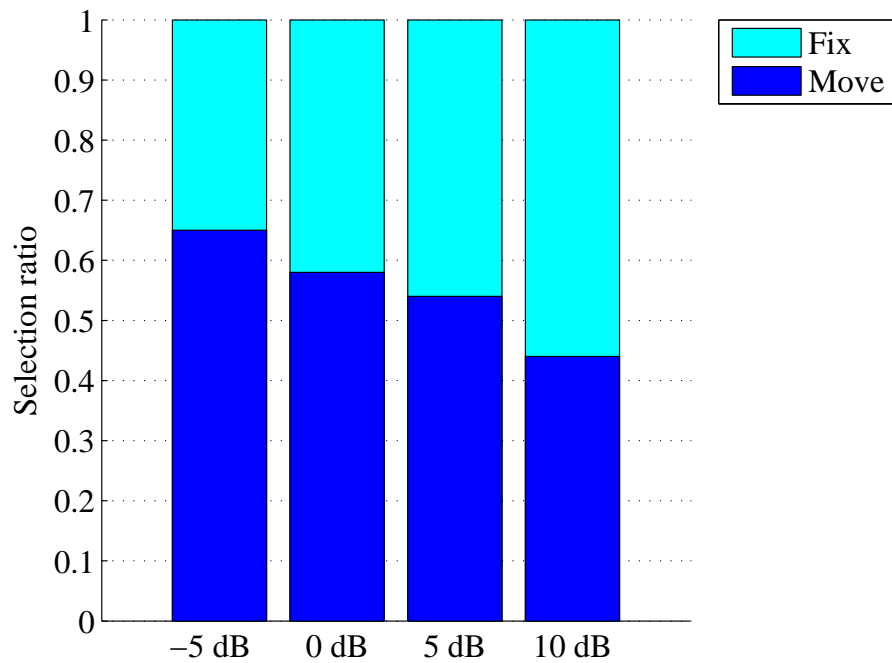


図 4.6: 実験 2-2 : SNR ごとの平均判断比率

ベルが相対的に高い場合ほど、頭部運動に感応する背景音を選択する比率が高くなる傾向が見られた。これは、背景音のレベルを高め設定するほど、頭部運動に感応しない背景音が不自然に感じられるためと考えられる。

さらに、実験後の内省報告によると、「実験 2-1 よりも違和感を感じなかった」、「対象音が自然に聞こえた」という意見が多かった。このことから、一部の方向の音圧レベルが高い背景音を付加した信号を提示した場合、全方向から同程度の音圧レベルで音が到来する背景音を付加した信号よりも、高い忠実性が得られる可能性が示唆された。

ここで、内省報告によると、被験者は実験刺激の違いに気が付いていなかったことがわかった。すなわち、提示した刺激 A と刺激 B について「A と B が違うのはわかるが、何が違うのかわからない」といった感想が大多数であった。背景音が動いていたことに気がついたのは被験者 9 名中 1 名であり、かつこの 1 名の判断比率は、他の 8 名のものと大きな差はなかった。したがって、被験者は無意識の処理によって忠実性を判断したものと考えられるが、何を原因として忠実性が向上したのかを今回の実験結果からで判断することはできず、改めて検証する必要があると考えられる。

4.4 実験 2-3：背景音源の音圧レベル分布をパラメータとした検証

実験 2-2 の結果から、Surround 背景音を構成する背景音源に振幅の差を与えた出力信号を頭部運動に感応させた信号は、同じ信号を頭部運動に感応させなかった場合に比べて有意に忠実性が高いことが確かめられた。しかし、何を原因として忠実性が向上したのかを実験結果からで判断することはできなかった。考えられる原因としては、頭部運動する音源が増えたためや、頭部運動に感応する背景音を自然に感じる以上に頭部運動に感応しない背景音が不自然に感じられたためなどが考えられる。

そこで、背景音が常に頭部運動に感応している状況で、背景音源の音圧レベルの分布をパラメータとした検証を行うこととした。

実験刺激の作成

式 (4.1) で定義される、2 次元 Gaussian 窓を用いる。ここで θ は水平角、 ϕ は仰角を表す。

$$w(\theta, \phi) = \exp\left(-\frac{\theta^2 + \phi^2}{\sigma^2}\right) \quad (4.1)$$

この σ^2 の値を変えることで、背景音の振幅に変化を与えることとする。

一例として、 $\sigma^2 = 0$ 、 $\sigma^2 = 8$ 、 $\sigma^2 = 64$ 、 $\sigma^2 = 512$ 、 $\sigma^2 = \infty$ 、における 2 次元 Gaussian 窓をそれぞれ図 4.7(a)、図 4.7(b)、図 4.7(c)、図 4.7(d)、図 4.7(e) に示す。なお、 $\sigma^2 = 0$ の場合、式 (4.1) は定義できないため、厳密には $\sigma^2 = \lim_{x \rightarrow +0} x$ であるが、簡略化のため上のように表記する。

なお、頭部中心位置における背景音のエネルギーの総和が一定になるように調整した。

実験方法

実験装置は前実験と同様のものを用いる。本実験において、SNR は $SNR = 10$ dB に統一した。対象音には、前実験までと同様に 4 種類の音源を使用した。背景音は、式 (4.1) の σ^2 の値を要因とし、 $\sigma^2 = 0$ 、 $\sigma^2 = 8$ 、 $\sigma^2 = 64$ 、 $\sigma^2 = 512$ 、 $\sigma^2 = \infty$ の 5 条件とした。以降、文中では各条件を A_0 条件、 A_8 条件、 A_{64} 条件、 A_{512} 条件、 A_∞ 条件と記述する。

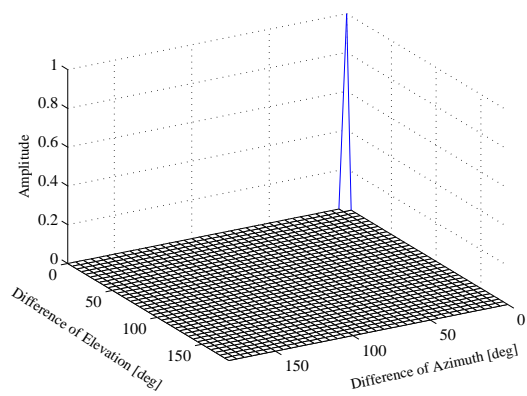
したがって、 A_0 条件は背景音を点音源とした条件で、対象音と背景音の二つの点音源が頭部運動に感応する条件であり、 A_∞ 条件は実験 2-1 の Move 条件と同じ刺激となる。

本実験においては、すべての提示音を頭部運動に感応させている。実験に参加した被験者は正常な聴力を持つ成人男女 11 名である。また、カウンターバランスを考慮して、提示順を逆にした場合についても実験を行った。したがって、一つの背景音対について、8 回の判断が成されることになり、本実験の総試行数は、聴取者一人当たり、計 ${}_5C_2 \times 8 = 80$ 試行となる。

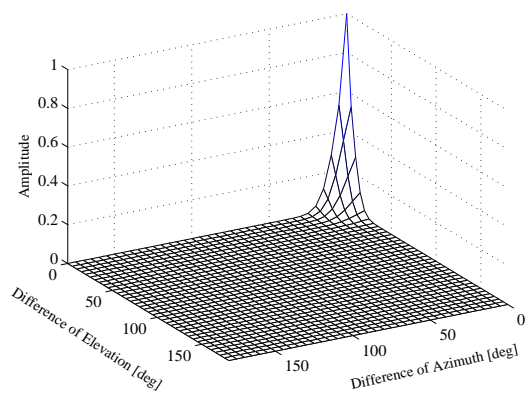
音圧校正は A_∞ 条件を基準として、 $L_{Aeq} = 60$ dB となるように設定した。

結果

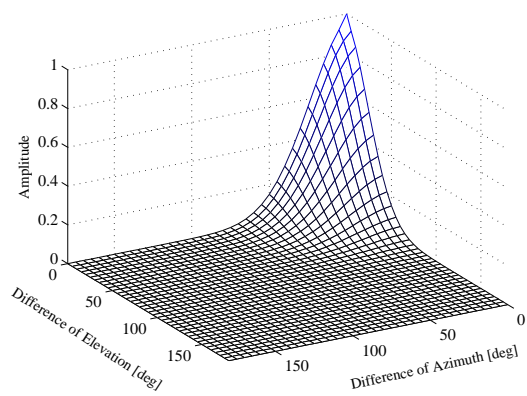
実験結果を図 4.8 に示す。図 4.8 は実験 2-3 における、参加者全員の各レンダリング手法に対する平均の心理距離尺度値を表わしている。分散分析の結果、条件の主効果が有意であった ($F(4, 54) = 22.335, p < .001$)。多重比較 (Bonferroni の不等式に基づく多重比較法) を行った結果、 A_{64} 条件は、 A_0 条件、 A_8 条件、及び A_∞ 条件よりも有意に値が高かった。 A_{512} 条件は、 A_0 条件、及び A_8 条件よりも有意に値が高かった。 A_∞ 条件は、 A_0 条件よりも有意に値が高かった。有意差のあった刺激の組み合わせを表 4.4 中に * 印で示す。



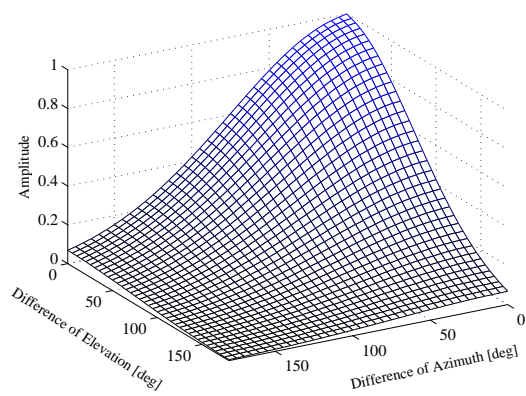
(a) $\sigma^2 = 0$



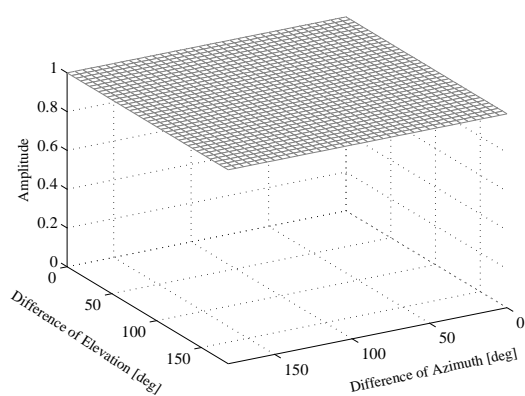
(b) $\sigma^2 = 8$



(c) $\sigma^2 = 64$



(d) $\sigma^2 = 512$



(e) $\sigma^2 = \infty$

図 4.7: 実験刺激の作成に用いた 2 次元 Gaussian 窓の比較

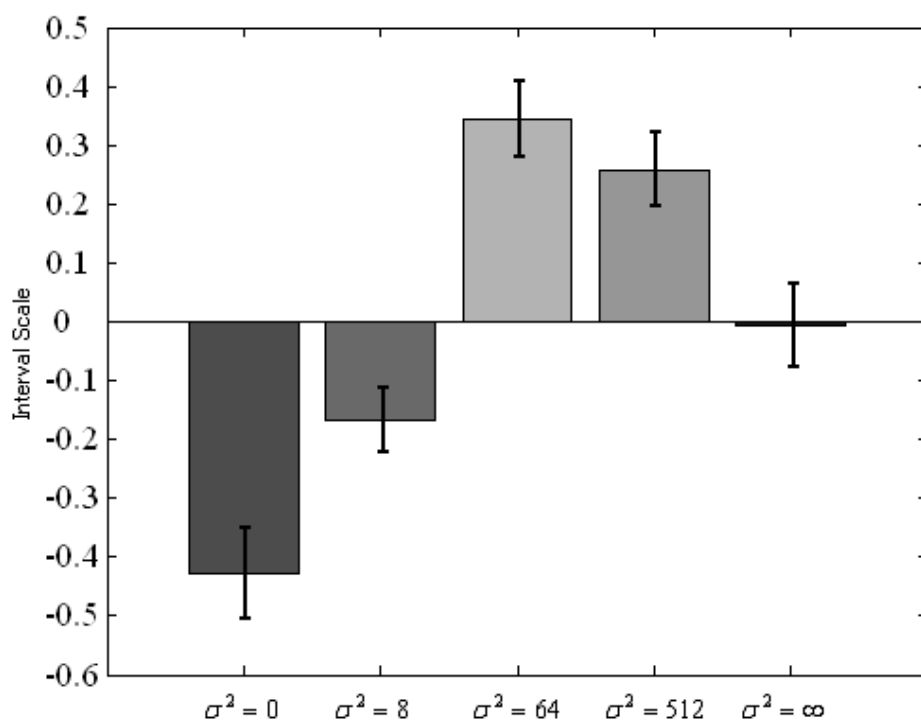


図 4.8: 実験 2-3 : 背景音源の振幅変化に対する心理距離尺度値 (参加者全体の平均)

表 4.4: 実験 2-3 : 有意差のあった刺激の組み合わせ

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-		*	*	*
	$\sigma^2 = 8$		-	*	*	
	$\sigma^2 = 64$	*	*	-		*
	$\sigma^2 = 512$	*	*		-	
	$\sigma^2 = \infty$	*		*		-

考察

背景音が点音源となる A_0 条件は、 A_{64} 条件、 A_{512} 条件、 A_∞ 条件よりも有意に低い値を示した。この結果から、忠実性を向上させるためには、全方向から音が聞こえるという状況が必要になると考えられる。

また、 A_{64} 条件、 A_{512} 条件は有意な差のないほぼ同程度の値であり、他の条件よりも比較的高い値を示した。特に A_{64} 条件は、各背景音源が同じ振幅である A_∞ 条件よりも有意に高い値を示した。この結果から、一部の振幅を高め、これを頭部運動に感応させることで、忠実性が向上することが示された。忠実性が向上した原因として、頭部運動に感応する音が増えたことで実空間に近い音環境の再現が可能になったことが考えられる。加えて、本実験では A_{64} 条件と A_{512} 条件を比較した場合に、被験者の選択傾向に偏りが見られた。例えば sub4, sub5, sub10, sub11 は A_{64} 条件を選択する比率が高かったのに対し、sub8, sub9 は A_{512} 条件を選択する比率が高かった。これは、忠実性を判定するための実空間を聴取者に自由に想定させたため、各被験者の想定した空間とこれらの条件を用いて作成された実験刺激が似ていたことに起因すると予想される。これは、 σ^2 のパラメータを適当に設定することで、ある程度の空間表現が可能になることを示唆している。

4.5 まとめ

本章では頭部運動と背景音が忠実性に与える影響の検証を行った。

前章の検討から、聴取者の頭部運動に出力が感応しない、静的な状況下においては、仮に背景音を加えたとしても、ある一定以上の忠実性を得ることはできないのではないかと予想されたため、動的状況を再現可能な頭部運動感応型聴覚ディスプレイを用いて検証を行った。また、本章で使用する背景音のベースには、Surround 背景音を使用した。

実験 2-1 において Surround 背景音を頭部運動に対応させたが、実験の結果、背景音を動かしたことによる差は見られなかった。原因として、Surround 背景音を構成する各背景音源の違いが、位相のみであったことが考えられた。

そこで、実験 2-2 では Surround 背景音を構成する背景音源に振幅差を与えた。その結果、頭部運動に感応させた背景音は、同背景音を感応させなかった場合に比べて有意に忠実性の向上が見られた。また、実験後の内省報告から、一部の方向の音圧レベルが高い背景音を付加した信号を提示した場合、全方向から同程度の音圧レベルで音が到来する背景音を付加した信号よりも、高い忠実性が得られる可能性が示唆された。

実験 2-3 で Surround 背景音を構成する背景音源の音圧レベルの分布を要因として実験を行った。背景音が点音源である場合、及び全方向から同程度の音圧レベルで音が到来する場合に比べて一部の方向の音圧レベルが高く、かつ広い範囲から音が到来する背景音を付加した信号を提示することで、高い忠実性が得られることが示された。

したがって、これらの三つの実験結果から、

1. 背景音が頭部運動に感応すること
2. 頭部周囲の全方向から擬似的に音が到来すること
3. 少なくとも一方向の音圧レベルが高いこと

以上三つの条件を満たす音源を用いれば、有意に音場の忠実性を向上させることが可能となると言える。

第5章 結論

現在，我が国の基盤産業としての地位を占めつつあるコンテンツ産業において，3次元空間のレンダリング技術を利用した高臨場感型コンテンツは極めて重要な枠割を果たすと考えられる．しかし，3次元空間のレンダリングは，視覚効果をターゲットにした画像処理については，盛んに技術開発やツールの開発が行われているのに対し，音響技術に関しては，それらが十分であるとは言えない．そこで，本研究では，忠実度の高いシステムの実現を目的として，3次元音響技術に関する検討を行った．また，本研究では，1つの音オブジェクトに対し1つのスレッドが生成されるというマルチスレッドの構造をとることや，Microsoft社のMicrosoft Windows上のDLLとして構築されていることなど，他の聴覚ディスプレイには見られない柔軟な構造を持つ，豊田らの開発したSiFASoへの実装を最終的な目的として検討を行うこととした．

第2章では，回折音場の最適レンダリング手法について検討を行った．現状のSiFASoは，自由空間及び直方体の室内空間しか模擬することができないため，曲がり角や障害物などの存在する空間をレンダリングすることが不可能であった，これらの環境をレンダリング可能にするためには，回折音のレンダリングが必須となる．回折音のレンダリング手法は，演算可能な環境が柔軟であること，及び演算が高速であることが条件として挙げられる．これらの点から，本研究ではCalamia-Svenssonの手法を採用することとした．Calamiaらの手法は，音源と聴取者の位置に制限がないため，柔軟性に優れ，演算も非常に高速であった．また，そのCalamiaらが回折音の演算部分に使用しているSvenssonらの手法は，有限長の障壁や，多段回折音に対応可能であり，拡張性に優れている．この手法を回折音処理クラス“CVADDif”，回折音付加スレッドクラス“CVADDifThread”の2つに分けてSiFASoに実装し，DLLとして使用可能とした．

第3章では，背景音の付加手法に関する検討を行った．既存の聴覚ディスプレイシステムは対象とする音のみを出力する仕様となっているため，出力される音は，非常に静かな録音スタジオでコンサートを聴いているかのような非常に不自然な音となっていた．そのため，本研究では出力に背景音を付加することで忠実性の向上を図った．まず，背景音として使用する音の検討を行った．本研究では屋外空間を含め

た一般的な状況下に適用可能な背景音の作成を目的としているが、背景音スペクトルとして一般的に知られる Hoth のスペクトルや三浦らのスペクトルは室内において測定されたものである。そこで、屋外の背景音を含めた様々な背景音のサンプルをもとに理想的な背景音源についての検討を行った結果、ホワイトノイズを $S(f) \propto 1/f^{2.0}$ で与えられる周波数特性でフィルタリングすることで得られる信号を背景音源として使用する。この背景音源を出力信号に単純加算しただけでは、背景音が頭内定位してしまうという問題があったため、252 個の無相関な背景音源群に、それぞれ異なる角度の HRIR を畳み込むことで、頭外定位する背景音を作成した。次に背景音の付加効果を検証するために主観評価による実験を行った。その結果、提案する背景音源を出力音に単純加算するだけでも忠実性が向上することが示された。また、単純に加算しただけでは実空間においてバイノーラル録音した背景音には及ばないものの、頭外定位する背景音を使用すれば、実測の背景音と同程度の忠実性を与えることが可能となることが示された。

第 4 章では、頭部運動を考慮した背景音のレンダリングに関する検討を行った。前章の検討から、聴取者の頭部運動に出力が感応しない、静的な状況下においては、仮に背景音を加えたとしても、ある一定以上の忠実性を得ることはできないのではないかと予想されたため、動的状況を再現可能な頭部運動感応型聴覚ディスプレイを用いて検証を行った。また、本章における背景音のベースには、Surround 背景音を使用し、各種検討事項を聴取実験によって検証し、結果をもとに忠実性を高めるための背景音の条件について考察した。まず、前章で提案した Surround 背景音を頭部運動に対応させたが、背景音を動かしたことによる差は見られなかった。原因として、Surround 背景音を構成する各背景音源の違いが、位相のみであったことが考えられた。そこで、Surround 背景音を構成する背景音源に振幅差を与えることで、頭部運動に伴う背景音の変化の認識が容易になるように改良を加えた。その結果、頭部運動に感応させた背景音は、同背景音を感応させなかった場合に比べて有意に忠実性の向上が見られた。また、実験後の内省報告から、一部の方向の音圧レベルが高い背景音を付加した信号を提示した場合、全方向から同程度の音圧レベルで音が到来する背景音を付加した信号よりも、高い忠実性が得られる可能性が示唆された。続いて、Surround 背景音を構成する背景音源の音圧レベルの分布を要因として実験を行った。背景音が点音源である場合、及び全方向から同程度の音圧レベルで音が到来する場合に比べて一部の方向の音圧レベルが高く、かつ広い範囲から音が到来する背景音を付加した信号を提示することで、高い忠実性が得られることが示された。したがって、これらの実験結果から、(1) 頭部周囲の全方向から擬似的に音が到来する、(2) 少なくとも一方向の音圧レベルが高い、(3) 頭部運動に感応する、これら 3 条件を満たす背景音を付加することによって、より忠実性の高い音空間を生成することができると

結論付けられた。

なお、第 2 章においては、1 段階のみの回折音をレンダリングしたが、実際の環境を考えた場合、例えば板状の単純な障壁においても、耳に入る信号はエッジ部分で 2 段階以上の回折が発生しているため、多段回折処理が必要になる。また、本研究における回折音は、自由空間のみに対応したものであり、回折音に対する反射音、あるいは反射音に対する回折音などには対応していない。SiFASo にはすでに初期反射音、ドップラー効果などの音響レンダリングアルゴリズムが実装されていることを考えると、これらのレンダリングアルゴリズムと本研究において実装した回折音レンダリングを組み合わせたシステムを開発することで、より忠実度の高いシステムの実現が期待できる。また、第 3 章、第 4 章において、本研究では特定の空間情報を持たない普遍的な背景音を使用した。実際の音空間を考えた場合、背景音は聴取者のいる空間、及び聴取者位置に依存するものであると考えられ、真に忠実度の高いシステムを構築する場合、空間依存性を持つ背景音の実現は極めて重要な項目になると考えられる。このように、忠実度の高い 3 次元音響システムの実現という大目標においては、多くの課題が残されている。

これらの課題に関しては、計算機の性能向上によって、ある程度解消可能な問題であると考えられるが、例えば回折音に関しては 3 段階以上の回折を表現するためのアルゴリズム、背景音に関しては物理的には無限に存在する背景音源を表現する手法など、検討すべき事項は数多く存在する。この目標の実現を目指した将来の研究に、本研究の成果が少しでも役立てられることを念願し、本論文の結びとする

付 録 A 3 章の実験結果の詳細

実験 1-1 において，対となっている刺激音における，判断の基準となる刺激（以下，基準刺激，Reference Stimulus）に比べて，判断比率を求める対象となる背景音条件（以下，対象刺激，Objective Stimulus）のほうがリアリティが高いと判断した比率を，各被験者について，表 A.1 から表 A.8 に示す．この判断比率を心理尺度値に変換し，各刺激音に対し心理距離尺度値を算出した．それを各被験者と平均値についてまとめたものを表 A.9 に示す．また，表 A.10 は記述統計量，表 A.11 は分散分析の結果，表 A.12 は Bonferroni の不等式に基づく多重比較の結果を示している．

A.1 実験 1-1

実験 1-1 において，Sub4 の心理距離尺度が一部発散したため，データから除いて処理を行っている．

表 A.1: sub1 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.31	0.44
	Normal	0.69	-	0.31
	Real	0.56	0.69	-

表 A.2: sub2 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.25	0.19
	Normal	0.75	-	0.5
	Real	0.81	0.5	-

表 A.3: sub3 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.44	0.38
	Normal	0.56	-	0.5
	Real	0.63	0.5	-

表 A.4: sub4 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0	0.06
	Normal	1	-	0.13
	Real	0.94	0.88	-

表 A.5: sub5 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.38	0.25
	Normal	0.63	-	0.25
	Real	0.75	0.75	-

表 A.6: sub6 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.5	0.31
	Normal	0.5	-	0.44
	Real	0.69	0.56	-

表 A.7: sub7 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.31	0.25
	Normal	0.69	-	0.13
	Real	0.75	0.88	-

表 A.8: sub8 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		None	Normal	Real
Reference stimulus	None	-	0.5	0.13
	Normal	0.5	-	0.38
	Real	0.88	0.63	-

表 A.9: 全被験者の心理距離尺度値

		Rendering Method		
		None	Normal	Real
sub1		-0.323	0.000	0.320
sub2		-0.781	0.337	0.444
sub3		-0.238	0.079	0.159
sub5		-0.497	-0.178	0.674
sub6		-0.244	-0.079	0.323
sub7		-0.582	-0.331	0.912
sub8		-0.575	-0.159	0.734
Average		-0.463	-0.047	0.509

表 A.10: 記述統計量

	度数	平均値	標準偏差	標準誤差	95 % 信頼区間		最小値	最大値
					下限	上限		
None	7	-0.4628	0.20282	0.07666	-0.6504	-0.2752	-0.78	-0.24
Normal	7	-0.0473	0.21484	0.08120	-0.2459	0.1514	-0.33	0.34
Real	7	0.5101	0.26978	0.10197	-0.2605	0.7596	-0.16	0.91
合計	21	0.0000	0.46349	0.10114	-0.2110	0.2110	-0.78	0.91

表 A.11: 分散分析表

	平方和	自由度	平均平方	F 値	有意確率
グループ間	3.336	2	1.668	31.261	0.000
グループ内	0.960	18	0.053		
合計	4.296	20			

表 A.12: 多重比較 (Bonferroni の不等式に基づく多重比較法)

条件 (I)	条件 (J)	平均値の差 ($I - J$)	標準誤差	有意確率	95 % 信頼区間	
					下限	上限
None	Normal	-0.41554*	0.12347	0.010	-0.7414	-0.0897
	Real	-0.97285*	0.12347	0.000	-1.2987	-0.6470
Normal	None	0.41554*	0.12347	0.010	0.0897	0.7414
	Real	-0.55731*	0.12347	0.001	-0.8832	-0.2315
Real	None	0.97285*	0.12347	0.000	0.6470	1.2987
	Normal	0.5571*	0.12347	0.001	0.2315	0.8832

A.2 実験 1-2

実験 1-2 における，各被験者の判断比率を表 A.13 から表 A.20 に示し，心理距離尺度値の平均値を表 A.21 に示す．表 A.22 は記述統計量，表 A.23 は分散分析の結果，表 A.24 は Bonferroni の不等式に基づく多重比較の結果を示す．

表 A.13: sub1 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.31	0.44
	Real	0.69	-	0.44
	Surround	0.56	0.56	-

表 A.14: sub2 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.44	0.38
	Real	0.56	-	0.44
	Surround	0.63	0.56	-

表 A.15: sub3 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.5	0.56
	Real	0.5	-	0.56
	Surround	0.44	0.44	-

表 A.16: sub4 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.13	0.38
	Real	0.88	-	0.81
	Surround	0.63	0.19	-

表 A.17: sub5 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.25	0.31
	Real	0.75	-	0.38
	Surround	0.69	0.63	-

表 A.18: sub6 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.44	0.44
	Real	0.56	-	0.69
	Surround	0.56	0.31	-

表 A.19: sub7 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.13	0.38
	Real	0.88	-	0.56
	Surround	0.63	0.44	-

表 A.20: sub8 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus		
		Normal	Real	Surround
Reference stimulus	Normal	-	0.38	0.25
	Real	0.63	-	0.5
	Surround	0.75	0.5	-

表 A.21: 実験 1-2 における全被験者の心理距離尺度値

	Rendering Method		
	Normal	Real	Surround
sub1	-0.323	0.166	0.157
sub2	-0.238	0.000	0.238
sub3	0.079	0.079	-0.157
sub4	-0.734	1.019	-0.284
sub5	-0.582	0.178	0.404
sub6	-0.157	0.323	-0.166
sub7	-0.734	0.654	0.081
sub8	-0.497	0.159	0.337
Average	-0.398	0.322	0.076

表 A.22: 記述統計量

	度数	平均値	標準偏差	標準誤差	95 % 信頼区間		最小値	最大値
					下限	上限		
Normal	8	-0.3984	0.28940	0.10232	-0.6403	-0.1564	-0.73	0.08
Real	8	0.3222	0.34437	0.12175	0.0343	0.6101	0.00	1.02
Surround	8	0.762	0.25387	0.08976	-0.1360	0.2884	-0.28	0.40
合計	24	0.0000	0.41775	0.08527	-0.1764	0.1764	-0.73	1.02

表 A.23: 分散分析表

	平方和	自由度	平均平方	F 値	有意確率
グループ間	2.146	2	1.073	12.067	0.000
グループ内	1.868	21	0.089		
合計	4.014	23			

表 A.24: 多重比較 (Bonferroni の不等式に基づく多重比較法)

条件 (I)	条件 (J)	平均値の差 ($I - J$)	標準誤差	有意確率	95 % 信頼区間	
					下限	上限
Normal	Real	-0.72051*	0.14911	0.000	-1.1084	-0.3326
	Surround	-0.47456*	0.14911	0.013	-0.8624	-0.0867
Real	Normal	0.72951*	0.14911	0.000	0.3326	1.1084
	Surround	0.24596	0.14911	0.342	-0.1419	0.6338
Surround	Normal	0.47456*	0.14911	0.013	0.867	0.8624
	Real	-0.24596	0.14911	0.342	-0.6338	0.1419

付 録 B 4 章の実験 2-3 結果の詳細

実験 2-3 において，対となっている刺激音における，判断の基準となる刺激（以下，基準刺激，Reference Stimulus）に比べて，判断比率を求める対象となる背景音条件（以下，対象刺激，Objective Stimulus）のほうがリアリティが高いと判断した比率を，各被験者について，表 B.1 から表 B.11 に示す．この判断比率を心理尺度値に変換し，各刺激音に対し心理距離尺度値を算出した．それを各被験者と平均値についてまとめたものを表 B.12 に示す．また，表 B.13 は記述統計量，表 B.14 は分散分析の結果，表 B.15 は Bonferroni の不等式に基づく多重比較の結果を示している．

表 B.1: sub1 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.25	0.38	0.38	0.38
	$\sigma^2 = 8$	0.75	-	0.50	0.25	0.50
	$\sigma^2 = 64$	0.63	0.50	-	0.50	0.63
	$\sigma^2 = 512$	0.63	0.75	0.50	-	0.38
	$\sigma^2 = \infty$	0.63	0.50	0.38	0.63	-

表 B.2: sub2 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.50	0.38	0.25	0.63
	$\sigma^2 = 8$	0.50	-	0.13	0.38	0.50
	$\sigma^2 = 64$	0.63	0.88	-	0.38	0.63
	$\sigma^2 = 512$	0.75	0.63	0.63	-	0.63
	$\sigma^2 = \infty$	0.38	0.50	0.38	0.38	-

表 B.3: sub3 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.50	0.38	0.25	0.38
	$\sigma^2 = 8$	0.50	-	0.63	0.38	0.50
	$\sigma^2 = 64$	0.63	0.38	-	0.50	0.88
	$\sigma^2 = 512$	0.75	0.63	0.50	-	0.63
	$\sigma^2 = \infty$	0.63	0.50	0.13	0.38	-

表 B.4: sub4 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.25	0.13	0.38	0.25
	$\sigma^2 = 8$	0.75	-	0.38	0.50	0.50
	$\sigma^2 = 64$	0.88	0.63	-	0.63	0.38
	$\sigma^2 = 512$	0.63	0.50	0.38	-	0.63
	$\sigma^2 = \infty$	0.75	0.50	0.63	0.38	-

表 B.5: sub5 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.50	0.50	0.38	0.38
	$\sigma^2 = 8$	0.50	-	0.38	0.50	0.63
	$\sigma^2 = 64$	0.50	0.63	-	0.88	0.50
	$\sigma^2 = 512$	0.63	0.50	0.13	-	0.63
	$\sigma^2 = \infty$	0.63	0.38	0.50	0.38	-

表 B.6: sub6 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.25	0.50	0.25	0.38
	$\sigma^2 = 8$	0.75	-	0.38	0.25	0.25
	$\sigma^2 = 64$	0.50	0.63	-	0.38	0.75
	$\sigma^2 = 512$	0.38	0.75	0.63	-	0.50
	$\sigma^2 = \infty$	0.63	0.75	0.25	0.50	-

表 B.7: sub7 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.63	0.13	0.38	0.25
	$\sigma^2 = 8$	0.38	-	0.50	0.25	0.50
	$\sigma^2 = 64$	0.88	0.50	-	0.50	0.63
	$\sigma^2 = 512$	0.63	0.75	0.50	-	0.50
	$\sigma^2 = \infty$	0.75	0.50	0.38	0.50	-

表 B.8: sub8 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.38	0.13	0.13	0.13
	$\sigma^2 = 8$	0.63	-	0.50	0.13	0.13
	$\sigma^2 = 64$	0.88	0.50	-	0.38	0.63
	$\sigma^2 = 512$	0.88	0.88	0.63	-	0.50
	$\sigma^2 = \infty$	0.88	0.88	0.38	0.50	-

表 B.9: sub9 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.38	0.50	0.25	0.38
	$\sigma^2 = 8$	0.63	-	0.38	0.25	0.50
	$\sigma^2 = 64$	0.50	0.63	-	0.38	0.50
	$\sigma^2 = 512$	0.75	0.75	0.63	-	0.50
	$\sigma^2 = \infty$	0.63	0.50	0.50	0.50	-

表 B.10: sub10 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.25	0.25	0.13	0.38
	$\sigma^2 = 8$	0.75	-	0.25	0.13	0.75
	$\sigma^2 = 64$	0.75	0.75	-	0.75	0.88
	$\sigma^2 = 512$	0.88	0.88	0.25	-	0.38
	$\sigma^2 = \infty$	0.63	0.25	0.13	0.63	-

表 B.11: sub11 が基準刺激に対して対象刺激がリアルであると判断した比率

		Objective stimulus				
		$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
Reference stimulus	$\sigma^2 = 0$	-	0.38	0.38	0.25	0.50
	$\sigma^2 = 8$	0.63	-	0.13	0.38	0.50
	$\sigma^2 = 64$	0.63	0.88	-	0.63	0.75
	$\sigma^2 = 512$	0.75	0.63	0.38	-	0.50
	$\sigma^2 = \infty$	0.50	0.50	0.25	0.50	-

表 B.12: 全被験者の心理距離尺度値

	Rendering Method				
	$\sigma^2 = 0$	$\sigma^2 = 8$	$\sigma^2 = 64$	$\sigma^2 = 512$	$\sigma^2 = \infty$
sub1	-0.408	0.000	0.159	0.169	0.080
sub2	-0.169	-0.367	0.367	0.408	-0.239
sub3	-0.328	0.000	0.288	0.328	-0.288
sub4	-0.704	0.089	0.367	0.080	0.169
sub5	-0.159	0.000	0.367	-0.128	-0.080
sub6	-0.169	-0.248	0.169	0.169	0.080
sub7	-0.456	-0.248	0.367	0.248	0.089
sub8	-0.942	-0.496	0.288	0.655	0.496
sub9	-0.328	-0.169	0.000	0.417	0.080
sub10	-0.704	-0.119	0.793	0.327	-0.297
sub11	-0.328	-0.288	0.616	0.169	-0.169
Average	-0.427	-0.168	0.344	0.258	-0.007

表 B.13: 記述統計量

	度数	平均値	標準偏差	標準誤差	95 % 信頼区間		最小値	最大値
					下限	上限		
$\sigma^2 = 0$	11	-0.4269	0.25603	0.07720	-0.5989	-0.2549	-0.94	-0.16
$\sigma^2 = 8$	11	-0.1678	0.18090	0.05454	-0.2893	-0.0462	-0.50	0.09
$\sigma^2 = 64$	11	0.3437	0.21621	0.06519	0.1985	0.4890	0.00	0.79
$\sigma^2 = 512$	11	0.2582	0.20489	0.06178	0.1205	0.3958	-0.13	0.65
$\sigma^2 = \infty$	11	-0.0072	0.23729	0.07155	-0.1667	0.1522	-0.30	0.50
合計	55	0.0000	0.35435	0.04778	-0.0958	0.0958	-0.94	0.79

表 B.14: 分散分析表

	平方和	自由度	平均平方	F 値	有意確率
グループ間	4.347	4	1.087	22.335	0.000
グループ内	2.433	50	0.049		
合計	6.781	54			

表 B.15: 多重比較 (Bonferroni の不等式に基づく多重比較法)

条件 (I)	条件 (J)	平均値の差 ($I - J$)	標準誤差	有意確率	95 % 信頼区間	
					下限	上限
$\sigma^2 = 0$	$\sigma^2 = 8$	-0.25910	0.09406	0.082	-0.5354	0.0172
	$\sigma^2 = 64$	-0.77061*	0.09406	0.000	-1.0469	-0.4944
	$\sigma^2 = 512$	-0.68503*	0.09406	0.000	-0.9613	-0.4088
	$\sigma^2 = \infty$	-0.41963*	0.09406	0.000	-0.6959	-0.1434
$\sigma^2 = 8$	$\sigma^2 = 0$	0.25910	0.09406	0.082	-0.0172	0.5354
	$\sigma^2 = 64$	-0.51151*	0.09406	0.000	-0.7878	-0.2353
	$\sigma^2 = 512$	-0.42593*	0.09406	0.000	-0.7022	-0.1497
	$\sigma^2 = \infty$	-0.16053	0.09406	0.941	-0.4368	0.1157
$\sigma^2 = 64$	$\sigma^2 = 0$	0.77061*	0.09406	0.000	0.4944	1.0469
	$\sigma^2 = 8$	0.51151*	0.09406	0.000	0.2353	0.7878
	$\sigma^2 = 512$	0.08558	0.09406	1.000	-0.1907	0.3618
	$\sigma^2 = \infty$	0.35098*	0.09406	0.005	0.0747	0.6272
$\sigma^2 = 512$	$\sigma^2 = 0$	0.68503*	0.09406	0.000	0.4088	0.9613
	$\sigma^2 = 8$	0.42593*	0.09406	0.000	0.1497	0.7022
	$\sigma^2 = 64$	-0.08558	0.09406	1.000	-0.3618	0.1907
	$\sigma^2 = \infty$	0.26540	0.09406	0.068	-0.0109	0.5417
$\sigma^2 = \infty$	$\sigma^2 = 0$	0.41963*	0.09406	0.000	0.1434	0.6959
	$\sigma^2 = 8$	0.16053	0.09406	0.941	-0.1157	0.4368
	$\sigma^2 = 64$	-0.35098*	0.09406	0.005	-0.6272	-0.0747
	$\sigma^2 = 512$	-0.26540	0.09406	0.068	-0.5417	0.0109

謝辞

本研究を行うにあたり，東北大学電気通信研究所鈴木陽一教授には研究の機会を賜り，研究を進めていく上で，多忙な身でありながらもご指導いただいた．また，常に真摯な態度で接してくださり，研究面のみならず，日常生活においても数々のご指導や，様々な知見をいただいた．ここに深く感謝申し上げます．

東北大学情報シナジーセンター木下哲男教授，同吉澤誠教授には，本論文の審査員として，有益な助言をいただいた．ここに深く感謝申し上げます．

東北大学電気通信研究所岩谷幸雄准教授には，本研究の直接なご指導をいただき，本論文の最初から最後まで研究の細部にわたる有益なご指導，ご助言をいただいた．また，日常生活においても未熟な筆者と視点をあわせて接してくださり，研究に行き詰った際に多くの励ましをいただいた．ここに深く感謝申し上げます．

東北大学大学院工学研究科牧野正三教授，伊藤彰則准教授，助教伊藤仁博士，東北大学大学院工学研究科塩入諭教授，栗木一郎准教授，助教松宮一道博士には，大学院ゼミの場において，数多くの有益な助言をいただいた．ここに深く感謝申し上げます．

東北大学電気通信研究所助教坂本修一博士，同宮内良太博士には，本研究を進めていく上で，数々の有益なご指摘，助言を頂いた．また，研究活動を進めていく上で研究環境の整備をしていただいた．ここに深く感謝申し上げます．東北大学電気通信研究所齋藤史孝技官には，音響測定の方法など懇切丁寧にご指導いただいた．ここに深く感謝申し上げます．東北大学電気通信研究所やわらかい情報システム研究センター助教矢入聡博士には本研究における有益な助言をいただき，また研究室の先輩として，様々な相談に乗っていただいた．ここに深く感謝申し上げます．東北大学工学研究科 COE フェロー大谷真博士には，建築音響，音響プログラミングに関する様々な助言をいただき，本研究における直接のご指導をいただいた．ここに深く感謝申し上げます．東北福祉大学情報福祉学科講大内誠准教授には，多忙な時間をぬって，様々な聴覚ディスプレイや SiFASo に関する様々な知識をいただいた．ここに深く感謝申し上げます．東北大学電気通信研究所非常勤研究員小林まおり博士，同共同研究員寺本渉博士には研究室ゼミの場で数々の有益なご指摘，ご助言をいただき，実験の統計手法に関する様々な助言をいただいた．ここに深く感謝申し上げます．研究支援者の宮内千夏氏にはプログラミングに明るくなかった筆者にもわかりやすく教えていただいた．

ここに深く感謝申し上げます。

東北大学鈴木研究室の先輩である東北大学大学院情報科学研究科博士課程後期 3 年 魏浩石氏，岡本拓磨氏，同 1 年浅川香氏には，本論文における数多くのご助言を頂いただけでなく，日常生活においても様々な相談に乗っていただいた。また，同 3 年 姜大基氏には，同じ空間音響グループとして，測定等の録音作業時に多大なるご尽力をいただいた。ここに深く感謝申し上げます。同輩の半田浩規氏には，共に修士論文に取り組む仲間として，研究に日常生活にと，忙しい中多くの協力と励ましをいただいた。苦勞しながらも，有意義な 3 年間の研究生生活が過ごせたのは，氏の存在あつてのことである。ここに深く感謝申し上げます。東北大学大学院情報科学研究科博士課程前期 1 年増山友介氏，藤井真治氏には，同じく空間音響グループとして，実験を行なううえでの準備や，測定に多大なる協力をいただいた。両氏に加え，小玉純一氏，佐藤渉氏，吉田和博氏，東北大学工学部情報工学科 4 年池田善敬氏，古根史雅氏，曲谷地哲氏，研究生ハン・チョル・ス氏には，忙しい中，被験者として実験に参加していただいた。ここに深く感謝申し上げます。

研究室事務補佐員の小野寺美紀氏，本間理恵氏には，研究に専念できるように様々なご配慮をいただいた。ここに深く感謝申し上げます。

本研究は，ここに記載することのできなかつた方々を含め，多くの方々の援助のもとに行われたものである。ご助力いただいた皆様に，重ねて深く感謝申し上げます。

最後に，私の学に対する意に理解を示し，長きにわたり支えてくれた家族に心より感謝する。

参考文献

- [1] 榎並和雅, 奥井誠人, 井ノ上直己: “NICT における超臨場感コミュニケーションの研究戦略,” *IEICE Technical Report*, EID2006-23, (2006).
- [2] Kimio Hamasaki, Toshiyuki Nishiguchi, Reiko Okumura, Yasushige Nakayama, and Akio Ando: “22.2 Multichannel Sound System for Ultra High-Definition TV,” AES 118th Convention, Barcelona, Spain, Convention paper 6406 (2005).
- [3] Adobe Photoshop:
<http://www.adobe.com/jp/products/photoshop/compare/>
- [4] Microsoft Direct3D:
<http://msdn.microsoft.com/ja-jp/library/cc351199.aspx>
- [5] H. Møller, M.F. Sørensen, C.B. Jenson and D.Hmmershøi: “Binaural technique: Do we need individual recording?,” *J. Audio Eng. Soc.*, **44**, 451-468 (1996).
- [6] H. Møller: “Fundamentals of binaural technology,” *Appl. Acoust.*, **36**, 171-218 (1992).
- [7] J. Blauert: “Spatial Hearing,” The MIT press, London (1983).
- [8] 渡邊貫治: “頭部伝達関数の補間と合成に関する研究,” 東北大学審査博士学位論文 (2004).
- [9] 豊田将志: “3次元音空間のレンダリング技法に関する研究,” 東北大学審査修士学位論文 (2005).
- [10] T. Lokki, J. Hiipakka, and L. Savioja: “A Framework for Evaluating Virtual Acoustic Environments,” AES 110TH Convention, Amsterdam, Netherland (2001).
- [11] SLAB:
<http://human-factors.arc.nasa.gov/SLAB/>

- [12] OpenAL:
<http://connect.creativelabs.com/openal/default.aspx>
- [13] 矢入聡, 岩谷幸雄, 鈴木陽一: “頭部運動感応型ソフトウェア聴覚ディスプレイの開発 ~ DSP を用いない高臨場感音空間合成システム ~,” 日本工業出版, 画像ラボ, 18(2), pp. 35-39 (2007).
- [14] E. M. Wenzel, J. D. Miller and J. S. Abel: “A software-based system for interactive spatial sound synthesis,” ICAD 2000, 6th Intl. Conf. on Aud. Disp (2000).
- [15] E. M. Wenzel, J. D. Miller and J. S. Able: “Sound lab: a real-time, software-based system for the study of spatial hearing,” 108th Conv. Aud. Eng. Soc (2000).
- [16] J. D. Miller: “Slab : a software-based real-time virtual acoustic environment rendering system,” ICAD 2001, 7th Intl. Conf. on Aud. Disp (2001).
- [17] J. D. Miller and E. M. Wenzel: “Recent developments in slab : a software-based system for interactive spatial sound synthesis,” ICAD 2002, 8th Intl. Conf. on Aud. Disp (2002).
- [18] L. Savioja, J. Huopaniemi, T. Lokki and R. Väänänen: “Creating interactive virtual acoustic environments,” *J. Audio Eng. Soc.*, 47(9), pp. 675-705 (1999).
- [19] T. Lokki: “Physically-Based Auralization - Design, Implementation, and Evaluation,” PhD thesis, Helsinki University of Technology (2002).
- [20] M. Ohuchi, Y. Iwaya, Y. Suzuki and T. Muneakata: “A game for visually impaired children with a 3-D virtual auditory display,” ICAD 2003, 9th Intl. Conf. on Aud. Disp (2003).
- [21] Tapio Lokki, and Hanna Järveläinen: “Subjective evaluation of auralization of physics-based room acoustics modeling,” Proceedings of the 2001 Intl. Conf. on Aud. Disp., Espoo, Finland, July 29-August 1 (2001).
- [22] 川井孝雄, 伊藤毅: “半無限障壁による回折音場,” 日本音響学会誌 32(5), pp.319-335 (1976).
- [23] 三浦康雄, 関口克明, 羽入敏樹, 坪山 睦: “有限音線積分法を用いた回折騒音のシミュレーション予測手法に関する基礎的検討: ウェッジ及び反無限障壁による回折音の計算,” Summaries of technical papers of Annual Meeting Architectural Institute of Japan (1999).

- [24] Wayne A. Kinney: “Scattering from a corrugated surface: Comparison between experiment, Helmholtz-Kirchhoff theory, and the facet-ensemble method,” *J. Acoust. Soc. Am.*, **73**, 183-194 (1983).
- [25] 古江嘉弘: “多面体による回折音場の計算法について (NDF 積分方程式に対する簡易補間式),” 日本建築学会大会学術公演梗概集 (1987).
- [26] 関喜一, 佐藤哲司: “3D サウンドを利用した聴覚空間認知訓練システムの開発,” 第 14 回視覚障害リハビリテーション研究発表大会 (2005).
- [27] M. Born, E. Wolf: “Principles of Optics,” Elsevier, pp. 370-556 (1975).
- [28] A. Sommerfeld: “Mathematische Theorie der Diffraction,” *Math. Ann.* **47** pp. 317-374 (1896).
- [29] H. S. Carslaw: “Diffraction of Waves by a Wedge of any Angle,” *Proc. London Math. Soc.*, pp. 291-306 (1920).
- [30] H. M. MacDonald: “The Addition-Theorem for the Bessel Functions,” *Proc. London Math. Soc.*, pp. 152-157 (1900).
- [31] 前川純一: “障壁の遮音設計に関する実験的研究,” *日本音響学会誌*, **18**, pp. 187-196 (1962).
- [32] 堀内俊宏, 東山三樹夫: “回折波を考慮した物陰音場のインパルス応答生成,” Technical Report of IEICE EA2000-18 (2000).
- [33] “岩波数学辞典第 3 版,” 岩波書店, pp.334-335.
- [34] M. A. Biot and I. Tolstoy: “Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction,” *J. Acoust. Soc. Am.*, **29**, 381-391 (1957).
- [35] H. Medwin, E. Childs and G. M. Jebsen: “Impulse studies of double diffraction: A discrete Huygens interpretation,” *J. Acoust. Soc. Am.*, **72**, 1005-1013 (1982).
- [36] U. Peter Svensson: “Numerical aspects of edge diffraction impulse responses,” *Acoust. Sci. & Tech.* **26**, 2 (2005).
- [37] S. Aoshima: “Compute-generated pulse signal applied for sound measurement,” *J. Acoust. Soc. Am.*, **69**, pp. 1484-1488 (1981).

- [38] U. Peter Svensson, and Roger I. Fred: “An analytic secondary source model of edge diffraction impulse responses,” *J. Acoust. Soc. Am.*, **106** (5), (1999).
- [39] T. Funkhouser, N. Tsingos, I. Carlbom: “A beam tracing method for interactive architectural acoustics,” *J. Acoust. Soc. Am.*, **115**(2), pp. 739-756 (2004).
- [40] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom: “Modeling acoustics in virtual environments using the uniform theory of diffraction,” Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01), pp. 545-552, Los Angeles, Calif, USA, August (2001).
- [41] Paul T. Calamia, and U. Peter Svensson: “Fast Time-Domain Edge-Diffraction Calculations for Interactive Acoustic Simulations,” *EURASIP Journal on Advances in Signal Processing* Volume 2007, Article ID 63560, 10 pages (2007).
- [42] 小貫輝義, 齋藤文孝, 鈴木陽一: “無指向性 32 面体スピーカの紹介”, 電気関係学会東北支部連合大会講演論文集, pp.220 (2005).
- [43] 吉田登美男, 岩崎俊一, 永井健三: “立体音の本質は何か 高級品質の因子分析 -,” 日本音響学会誌 **16**, pp.249-257 (1960).
- [44] 島田正治, 山崎芳男: “通信網における信号処理技術の一展開,” 日本音響学会誌 **47** (1991).
- [45] D. F. Hoth: “Room noise spectra at subscribers’s telephone locations,” *J. Acoust. Soc. Am.*, **12**, pp.499 (1941).
- [46] 三浦種敏, 山口善司, 鈴木明, 都丸喜成: “東京都内加入者室内騒音スペクトルについて,” 日本音響学会誌, **8**, pp.153-157 (1952).
- [47] 日本建築学会編: “建築と環境のサウンドライブラリ”, 技報堂出版 (2004).
- [48] K. Kawai, K. Fujimoto, T. Iwase, T. Sakuma, Y. Hidaka and H. Yasuoka: “Introduction of sound material in living environment 2004 (SMILE 2004),” A sound source database for educational and practical purposes, 4th Joint Meet. ASA/ASJ (Honolulu), *J. Acoust. Soc. Am.* **120**, pp.3070-3071 (2006).
- [49] ISO 3382:1997, “Measurement of the reverberation time of rooms with reference to other acoustical parameters”

- [50] M. Matsumoto and T. Nishimura: “Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator,” *ACM Trans. on Modeling and Computer Simulations* (1998).
- [51] 平原達也, 青山祐樹, 大谷真: “聴覚実験に用いるイヤホンの諸特性”, *日本音響学会講演論文集*, **107** pp. 25-30 (2007).
- [52] 永田靖ら: “統計的多重比較法の基礎,” *サイエンティスト社*, pp.81 (1997)
- [53] H. Wallach: “On sound localization,” *J. Acoust. Soc. Am.*, **10**, 270 (1939).
- [54] H. Wallach: “The role of head movements and vestibular and visual cues in sound localization,” *J. exp. psychol.*, **27**, 339 (1940).
- [55] J. Kawaura, Y. Suzuki, F. Asano and T. Sone: “Sound localization in head-phone reproduction by simulating transfer function from the sound source to the external ear,” *Acoust. Sci. & Tech.*, **12**, 203 (1991).
- [56] R. D. Sorkin, F. L. Wightman, D. S. Kistler and G. Elever: “An exploratory study on the use of movement-correlated cues in auditory head-up display,” *Hum. Factor.*, **31**, 161 (1989).
- [57] 佐々木寿幸: “仮想音の音像定位における聴取者の頭部運動の寄与に関する研究,” *東北大学審査修士学位論文* (2003).
- [58] M. Otani and T. Hirahara: “Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display,” *IEICE Trans. on Fundamentals*, **E91-A**, No.6, pp. 1320-1328 (2008).
- [59] W. R. Thurlow and P. S. Runge; “Effect of induced head movement in localization of direction of sound,” *J. Acoust. Soc. Am.*, **42**, 480 (1967).